Embedding Metrics into Ultrametrics and Graphs into Spanning Trees with Constant Average Distortion

Ittai Abraham* Yair Bartal† Ofer Neiman‡
Hebrew University

February 1, 2008

Abstract

This paper addresses the basic question of how well can a tree approximate distances of a metric space or a graph. Given a graph, the problem of constructing a spanning tree in a graph which strongly preserves distances in the graph is a fundamental problem in network design. We present scaling distortion embeddings where the distortion scales as a function of ϵ, with the guarantee that for each ϵ the distortion of a fraction $1-\epsilon$ of all pairs is bounded accordingly. Such a bound implies, in particular, that the *average distortion* and ℓ_q-distortions are small. Specifically, our embeddings have *constant* average distortion and $O(\sqrt{\log n})$ ℓ_2-distortion. This follows from the following results: we prove that any metric space embeds into an ultrametric with scaling distortion $O(\sqrt{1/\epsilon})$. For the graph setting we prove that any weighted graph contains a spanning tree with scaling distortion $O(\sqrt{1/\epsilon})$. These bounds are tight even for embedding in arbitrary trees. For probabilistic embedding into spanning trees we prove a scaling distortion of $\tilde{O}(\log^2(1/\epsilon))$, which implies *constant* ℓ_q-distortion for every fixed $q < \infty$.

1 Introduction

The problem of embedding general metric spaces into tree metrics with small distortion has been central to the modern theory of finite metric spaces. Such embeddings provide an efficient representation of the complex metric structure by a very simple metric. Moreover, the special class of ultrametrics (rooted trees with equal distances to the leaves) plays a special role in such embeddings [6, 9]. Such an embedding provides an even more structured representation of the space which has a hierarchical structure [6]. Probabilistic embedding into ultrametrics have led to algorithmic application for a wide range of problems (see [18]). An important problem in network design is to find a tree spanning the network, represented by a graph, which provides good approximation of the metric defined with the shortest path distances in the graph. Different notions have been suggested to quantify how well distances are preserved, e.g. routing trees and communication trees [23]. The papers [3, 12] study the problem of constructing a spanning tree with low average stretch, i.e., low average distortion over the edges of the tree. It is natural to define our measure of quality for the embedding to be its average distortion over all pairs, or alternatively the more strict measure of its ℓ_2-distortion. Such notions are very common in most practical studies of embeddings (see for example [16, 17, 4, 14, 21, 22]) . We recall the definitions from [2]: Given two metric spaces (X, d_X) and (Y, d_Y) an *injective* mapping $f : X \to Y$ is called an *embedding* of X into Y. An embedding is *non-contractive* if for any $u \neq v \in X$: $d_Y(f(u), f(v)) \geq d_X(u, v)$. For a non-contractive embedding let the distortion of the pair $\{u, v\}$ be $\mathrm{dist}_f(u, v) = \frac{d_Y(f(u), f(v))}{d_X(u, v)}$.

*email: ittaia@cs.huji.ac.il.
†email: yair@cs.huji.ac.il. Supported in part by a grant from the Israeli Science Foundation (195/02).
‡email: neiman@cs.huji.ac.il. Supported in part by a grant from the Israeli Science Foundation (195/02).

Definition 1 (ℓ_q-distortion). *For $1 \le q \le \infty$, define the* ℓ_q-distortion *of an embedding f as:*

$$\mathrm{dist}_q(f) = \|\mathrm{dist}_f(u,v)\|_q^{(\mathcal{U})} = \mathbb{E}[\mathrm{dist}_f(u,v)^q]^{1/q},$$

where the expectation is taken according to the uniform distribution $\mathcal{U}$ over $\binom{X}{2}$. The classic notion of distortion *is expressed by the ℓ_∞-distortion and the* average distortion *is expressed by the ℓ_1-distortion.*

The notion of average distortion is tightly related (see [2]) to that of embedding with scaling distortion [19, 1, 2].

Definition 2 (Partial/Scaling Embedding). *Given two metric spaces (X, d_X) and (Y, d_Y), a* partial embedding *is a pair (f, G), where f is a non-contractive embedding of X into Y, and $G \subseteq \binom{X}{2}$. The distortion of (f, G) is defined as: $\mathrm{dist}(f, G) = \sup_{\{u,v\}\in G} \mathrm{dist}_f(u, v)$. For $\epsilon \in [0, 1)$, a $(1-\epsilon)$-*partial embedding *is a partial embedding such that $|G| \ge (1-\epsilon)\binom{n}{2}$.*[1] *Given two metric spaces (X, d_X) and (Y, d_Y) and a function $\alpha : [0, 1) \to \mathbb{R}^+$, we say that an embedding $f : X \to Y$ has* scaling distortion *α if for any $\epsilon \in [0, 1)$, there is some set $G(\epsilon)$ such that $(f, G(\epsilon))$ is a $(1-\epsilon)$-partial embedding with distortion at most $\alpha(\epsilon)$.*

We prove the following theorems:

Theorem 1. *Any n-point metric space embeds into an ultrametric with scaling distortion $O(\sqrt{1/\epsilon})$. In particular, its ℓ_q-distortion is $O(1)$ for $1 \le q < 2$, $O(\sqrt{\log n})$ for $q = 2$, and $O(n^{1-2/q})$ for $2 < q \le \infty$.*

Theorem 2. *Any weighted graph of size n contains a spanning tree with scaling distortion $O(\sqrt{1/\epsilon})$. In particular, its ℓ_q-distortion is $O(1)$ for $1 \le q < 2$, $O(\sqrt{\log n})$ for $q = 2$, and $O(n^{1-2/q})$ for $2 < q \le \infty$.*

We show that the bounds in Theorems 1 and 2 are tight for the n-node cycle even for embeddings into arbitrary tree metrics. We also obtain an equivalent result for probabilistic embedding into spanning trees:

Theorem 3. *Any weighted graph of size n probabilistically embeds into a spanning tree with scaling distortion $\tilde{O}(\log^2 1/\epsilon)$. In particular, its ℓ_q-distortion is $O(1)$ for any fixed $1 \le q < \infty$*[2].

1.1 Related Work

Embedding metrics into trees and ultrametrics was introduced in the context of probabilistic embedding in [6]. Other related results on embedding into ultrametrics include work on metric Ramsey theory [9], multi-embeddings [11] and dimension reduction [10]. Embedding an arbitrary metric into a tree metric requires $\Omega(n)$ distortion in the worst case even for the metric of the n-cycle [20]. It is a simple fact [15, 9, 6] that any n-point metric embeds in an ultrametric with distortion $n-1$. However the known constructions are not scaling and have average distortion linear in n. The probabilistic embedding theorem [13, 8] (improving earlier results of [6, 7]) states that any n-point metric space probabilistically embeds into an ultrametric with distortion $O(\log n)$. This result has been the basis to many algorithmic applications (see [18]). This theorem implies the existence of a single ultrametric with average distortion $O(\log n)$ (a constructive version was given in [8]). This bound was later improved with the analysis of [1] as we discuss below. The study of partial embedding and scaling distortion was initiated by Kleinberg, Slivkins and Wexler [19], and later studied in [1, 2]. Abraham et. al [1] prove that any finite metric space probabilistically embeds in an ultrametric with scaling distortion $O(\log(1/\epsilon))$ implying constant average distortion. As mentioned above, since the distortion is bounded in expectation, this result implies the existence of a single ultrametric with constant average distortion, but does not bound the ℓ_2-distortion. In [2] we have studied in depth the notions of average distortion and ℓ_q-distortion and their relation to partial and scaling embeddings. Our main focus was the study of optimal scaling embeddings for embedding into L_p spaces. For embedding of

[1] Note that the embedding is *strictly* partial only if $\epsilon \ge 1/\binom{n}{2}$.

[2] Note that probabilistic embedding bounds on the ℓ_q-distortion *do not* imply an embedding into a single tree with the same bounds, with the exception of $q = 1$.

metrics into ultrametrics, we mentioned that *partial* embeddings exist with distortion $O(\sqrt{1/\epsilon})$ matching the lower bound from [1]. Theorem 1 significantly strengthens this result by providing an embedding with *scaling* distortion. That is, the bound holds for all values of $0 < \epsilon < 1$ *simultaneously* and therefore the embedding has bounded ℓ_q-distortion. It is a basic fact that the minimum spanning tree in an n-point weighted graph preserves the (shortest paths) metric associated with the graph up to a factor of $n-1$ at most. This bound is tight for the n-cycle. Here too, it is easy to see that the MST does not have scaling distortion, and may result in linear average distortion. Alon, Karp, Peleg and West [3] studied the problem of computing a spanning tree of a graph with small average stretch (over the edges of the graph). This can also be viewed as the dual of probabilistic embedding of the graph metric in spanning trees. Their work was recently significantly improved by Elkin, Emek, Speilman and Teng [12] who show that any weighted graph contains a spanning tree with average stretch $O(\log^2 n \log\log n)$. This result can also be rephrased in terms of the average distortion (but not the ℓ_2-distortion) over all pairs. For spanning trees, this paper gives the *first* construction with *constant* average distortion.

1.2 Discussion of Techniques

Theorem 1 uses partitioning techniques similar to those used in the context of the metric Ramsey problem [5, 9]. However, in our case we need to provide an argument for the existence of a partition which simultaneously satisfies multiple conditions, each for every possible value of ϵ. Theorem 2 builds on the technique above together with the Elkin et. al. [12] method to construct a spanning tree. A straightforward application of this approach loses an extra $O(\log n)$ factor and hence does not give a scaling distortion depending solely on ϵ. The loss in the Elkin et.al. approach stems from the need to bound the diameter in the recursive construction of the spanning tree. In each level of the construction we may alow only a very small increase as these get multiplied in the bound on the total blow up in the overall diameter. In their original work [12] the increase per level is $\Theta(1/\log n)$ which translates to the blow up in the distortion. In our case we show that the increase can exponentially decrease along the levels. This indeed guarantees a good blow up in the overall diameter but is awful in terms of the distortion. We apply a new technique for bounding the diameter which allows us to limit the number of levels involved. On the other hand it is clear that for every value of ϵ there is a limited number of levels for which the distortion requirement imposes new constraints. The proof then proceeds to carefully balance these different arguments. Theorem 3 uses essentially the same ideas together with the known probabilistic embedding methods (in fact, the proof of this theorem is somewhat less technically involved). The fact that these theorems are tight essentially follows from the results and techniques of [1, 2].

2 Preliminaries

Consider a finite metric space (X, d) and let $n = |X|$. For any point $x \in X$ and a subset $S \subseteq X$ let $d(x, S) = \min_{s \in S} d(x, s)$. The *diameter* of X is denoted $\mathrm{diam}(X) = \max_{x,y \in X} d(x, y)$. For a point $x \in X$ and $r \geq 0$, the ball at radius r around x is defined as $B_X(x, r) = \{z \in X | d(x, z) \leq r\}$. We omit the subscript X when it is clear form the context. Given $x \in X$ let $\mathrm{rad}_x(X) = \max_{y \in X} d(x, y)$. When a cluster X has a center $x \in X$ that is clear from the context we will omit the subscript and write $\mathrm{rad}(X)$ instead of $\mathrm{rad}_x(X)$. Given an edge-weighted graph $G = (X, E, \omega)$ with $\omega : E \to \mathbb{R}^+$, let (X, d) be the metric space induced from the graph in the usual manner - vertices are associated with points, distances between points correspond to shortest-path distances in G.

Definition 3. *An ultrametric U is a metric space (U, d_U) whose elements are the leaves of a rooted labelled tree T. Each $v \in T$ is associated a label $\Phi(v) \geq 0$ such that if $u \in T$ is a descendant of v then $\Phi(u) \leq \Phi(v)$ and $\Phi(u) = 0$ iff $u \in U$ is a leaf. The distance between leaves $x, y \in U$ is defined as $d_U(x, y) = \Phi(\mathrm{lca}(x, y))$ where $\mathrm{lca}(x, y)$ is the least common ancestor of x and y in T.*

3 Scaling embedding into an ultrametric

Theorem 4. *Any n-point metric space embeds into an ultrametric with scaling distortion $O(\sqrt{1/\epsilon})$. In particular, its ℓ_q-distortion is $O(1)$ for $1 \le q < 2$, $O(\sqrt{\log n})$ for $q = 2$, and $O(n^{1-2/q})$ for $2 < q \le \infty$.*

We give the proof for scaling distortion. The consequence of the bounds on the ℓ_q-distortion follows by a simple calculation. The proof is by induction on the size of X (the base case is where $|X| = 1$ and is trivial). Assume the claim is true for any metric space with less than n points. Let (X, d) be a metric space with $n = |X|$ and $\Delta = \mathrm{diam}(X)$. The ultrametric is defined in a standard manner by defining the labelled tree T whose leaf-set is X. The high level construction of T is as follows: find a partition P of X into X_1 and $X_2 = X \setminus X_1$, the root of T will be labelled Δ, and its children T_1, T_2 will be the trees formed recursively from the ultrametric trees of X_1 and X_2 respectively. Let $u \in X$ be such that $|B(u, \Delta/2)| \le n/2$ (such a point can always be found). For any $0 < \epsilon \le 1$ denote by $B_\epsilon(X)$ the total number of pairs $(x, y) \in X$ such that $d_T(x, y) > (150/\sqrt{\epsilon}) d_X(x, y)$. For a partition $P = (X_1; X_2)$ let $\hat{B}_\epsilon(P) = |\{(x, y) \mid x \in X_1 \wedge y \in X_2 \wedge d_X(x, y) \le (\sqrt{\epsilon}/150) \cdot \Delta\}|$.

Claim 1. *Let $\epsilon \in (0, 1]$ and let (X, d) be a metric space, if for any sub metric $X' \subseteq X$ there exists a partition $P = (X_1; X_2)$ be a partition of X' such that $\hat{B}_\epsilon(P) < \epsilon |X_1| \cdot |X_2|$ then $B_\epsilon(X) \le \epsilon \binom{|X|}{2}$.*

Proof. Let $P = (X_1; X_2)$ be a partition of X such that $\hat{B}_\epsilon(P) \le \epsilon |X_1| \cdot |X_2|$. By induction,

$$\begin{aligned}
B_\epsilon(X) &\le \hat{B}_\epsilon(P) + B_\epsilon(X_1) + B_\epsilon(X_2) \\
&\le \epsilon \left(\binom{|X_1|}{2} + \binom{|X_2|}{2} + |X_1| \cdot |X_2| \right) \\
&= \epsilon/2 \left(|X_1|^2 - |X_1| + |X_2|^2 - |X_2| + 2|X_1| \cdot |X_2| \right) \\
&= \epsilon/2 \left((|X_1| + |X_2|)(|X_1| + |X_2| - 1) \right) \\
&= \epsilon \binom{|X|}{2}.
\end{aligned}$$

$\square$

So it is sufficient to show that there exists a partition satisfying Claim 1 for all $\epsilon \in (0, 1]$ simultaneously.

Partition Algorithm. Let $\hat{\epsilon} = \max\{\epsilon \in (0, 1] \mid |B(u, \sqrt{\epsilon}\Delta/4)| \ge \epsilon n\}$. Observe that $1/n \le \hat{\epsilon} \le 1/2$ by the choice of u . Define the intervals $\hat{S} = [\sqrt{\hat{\epsilon}}\Delta/4, \sqrt{\hat{\epsilon}}\Delta/2]$, $S = [(\frac{1}{4} + \frac{1}{25})\sqrt{\hat{\epsilon}}\Delta, (\frac{1}{2} - \frac{1}{25})\sqrt{\hat{\epsilon}}\Delta]$, $s = \frac{17}{100}\sqrt{\hat{\epsilon}}\Delta$, and the shell $Q = \{w \mid d(u, w) \in \hat{S}\}$. We partition X by choosing some $r \in S$ such that $X_1 = B(u, r)$ and $X_2 = X \setminus X_1$. The following property will be used in several cases:

Claim 2. $|B(u, \sqrt{\hat{\epsilon}}\Delta/2)| \le 4\hat{\epsilon} n$.

Proof. There are two cases: If $\hat{\epsilon} \le 1/4$ then $|B(u, \sqrt{\hat{\epsilon}}\Delta/2)| = |B(u, \sqrt{4\hat{\epsilon}}\Delta/4)| \le 4\hat{\epsilon} n$ (otherwise contradiction to maximality of $\hat{\epsilon}$). Otherwise, $\hat{\epsilon} \in (1/4, 1]$. In such a case $|B(u, \sqrt{\hat{\epsilon}}\Delta/2)| \le |B(u, \Delta/2)| \le n/2 \le 2\hat{\epsilon} n$. $\square$

We will now show that some choice of $r \in S$ will produce a partition that satisfies Claim 1 for all $\epsilon \in (0, 32\hat{\epsilon}]$. For any $r \in S$ and $\epsilon \le 32\hat{\epsilon}$ let $S_r(\epsilon) = (r - \sqrt{\epsilon}\Delta/150, r + \sqrt{\epsilon}\Delta/150)$, $s(\epsilon) = \sqrt{\epsilon}\Delta/75$, and let $Q_r(\epsilon) = \{w \mid d(u, w) \in S_r(\epsilon)\}$. Notice that for any $r \in S$ and any $\epsilon \le 32\hat{\epsilon}$: $S_r(\epsilon) \subseteq \hat{S}$. Define that properly $A_r(\epsilon)$ holds if cutting at radius r is "good" for ϵ, formally: $A_r(\epsilon)$ iff $|Q_r(\epsilon)| < \sqrt{\epsilon \cdot \hat{\epsilon}/2} \cdot n$. For any $\epsilon \le 32\hat{\epsilon}$, note that in any partition to $X_1 = B(u, r)$, $X_2 = X \setminus X_1$ only pairs (x, y) such that $x, y \in Q_r(\epsilon)$ are distorted by more than $O(\sqrt{1/\epsilon})$. If property $A_r(\epsilon)$ holds then $\hat{B}_\epsilon(P) \le \epsilon \cdot \hat{\epsilon} n^2/2$. Since $\hat{\epsilon} n \le |X_1| \le n/2$ then $\epsilon \cdot \hat{\epsilon} n^2/2 \le \epsilon n/2 |X_1| \le \epsilon |X_1||X_2|$ so $A_r(\epsilon)$ implies Claim 1 for ϵ. Hence for $\epsilon \in (0, 32\hat{\epsilon}]$ the following is sufficient:

Claim 3. *There exists some $r \in S$ such that properly $A_r(\epsilon)$ holds for all $\epsilon \in (0, 32\hat{\epsilon}]$.*

Proof. The proof is based on the following iterative process that greedily deletes the "worst" interval in S. Initially, let $I_0 = S$, and $j = 1$:

1. If for all $r \in I_{j-1}$ and for all $\epsilon \le 32\hat{\epsilon}$ property $A_r(\epsilon)$ holds then set $t = j - 1$, stop the iterative process and output I_t.

2. Let $\mathcal{S}_j = \{S_r(\epsilon) \mid r \in I_{j-1}, \epsilon \le 32\hat{\epsilon}, \neg A_r(\epsilon)\}$. We greedily remove the interval $S \in \mathcal{S}_j$ that has maximal ϵ. Formally, let r_j, ϵ_j be parameters such that $S_{r_j}(\epsilon_j) \in \mathcal{S}_j$ and $\epsilon_j = \max\{\epsilon \mid \exists S_r(\epsilon) \in \mathcal{S}_j\}$.

3. Set $I_j = I_{j-1} \setminus S_{r_j}(\epsilon_j)$, set $j = j + 1$, and goto 1.

Let $\mathcal{Q} = \{Q_r(\epsilon)\}$ and note that $|\mathcal{Q}| = O(n^2)$ and it is easy to show that for every $j \in \{1, \ldots, t\}$, $Q' \in \mathcal{Q}$, the maximum of $\{\epsilon \mid S_r(\epsilon) \in \mathcal{S}_j, Q_r(\epsilon) = Q\}$ is obtained inside the set and can be found in $O(n^2)$ time.

We now argue that $I_t \neq \emptyset$ and hence such a value $r \in S$ can be found. Since for any $1 \le j < i \le t$, $s(\epsilon_j) \ge s(\epsilon_i)$ it follows that any $x \in Q$ appears in at most 2 "bad" intervals. From this and Claim 2:

$$\sum_{j=1}^{t} |Q_{r_j}(\epsilon_j)| \le 2|Q| \le 8\hat{\epsilon}n.$$

Recall that since $A_{r_j}(\epsilon_j)$ does not hold then for any $1 \le j \le t$: $|Q_{r_j}(\epsilon_j)| \ge \sqrt{\epsilon_j \cdot \hat{\epsilon}/2} \cdot n$ which implies that

$$\sum_{j=1}^{t} \sqrt{\epsilon_j} \le 12\sqrt{\hat{\epsilon}}.$$

On the other hand, by definition

$$\sum_{j=1}^{t} s(\epsilon_j) \le \sum_{j=1}^{t} \sqrt{\epsilon_j}\Delta/75 \le 12/75 \cdot \sqrt{\hat{\epsilon}}\Delta = 16/100 \cdot \sqrt{\hat{\epsilon}}\Delta.$$

Since $s = 17/100 \cdot \sqrt{\hat{\epsilon}}\Delta$ then indeed $I_t \neq \emptyset$ so any $r \in I_t$ satisfies the condition of the claim. $\square$

It remains to show that any choice of $r \in S$ will produce a partition that satisfies Claim 1 for all $\epsilon \in (32\hat{\epsilon}, 1]$.

Claim 4. *If $\epsilon \in (32\hat{\epsilon}, 1]$, $r \in S$ and $P = (B(u, r); X \setminus B(u, r))$ then $\hat{B}_\epsilon(P) < \epsilon|X_1| \cdot |X_2|$.*

Proof. Let $\epsilon \in (32\hat{\epsilon}, 1]$ and fix some $r \in S$. Only pairs (x, y) such that $x \in X_1$ and $y \in B(u, r + \sqrt{\epsilon}\Delta/16) \cap X_2$ can be distorted by more than $16\sqrt{1/\epsilon}$ and hence may be counted in $\hat{B}_\epsilon(P)$. Since $\sqrt{\hat{\epsilon}} \le \sqrt{\epsilon/2}/4$ and $r < \sqrt{\hat{\epsilon}}\Delta/2$ then $|B(u, r + \sqrt{\epsilon}\Delta/16)| \le |B(u, \sqrt{\epsilon/2}(\frac{1}{8} + \frac{1}{8})\Delta)| = |B(u, \sqrt{\epsilon/2}\Delta/4)| < \epsilon n/2$ by the maximality of $\hat{\epsilon}$. Since $|X_2| \ge n/2$ it follows that $\hat{B}_\epsilon(P) \le \epsilon|X_1| \cdot |X_2|$, as required. $\square$

Proof of Theorem 1. From Claim 3 and Claim 4, it follows that our partition scheme finds a cut $P = (X_1; X_2)$ such that $\hat{B}_\epsilon(P) < \epsilon|X_1| \cdot |X_2|$ for all ϵ. Hence when applying the partition scheme inductively, by Claim 1 the theorem follows. $\square$

4 Scaling Embedding into a Spanning Tree

Here we extended the techniques of the previous section, in conjunction with the constructions of [12] to achieve the following:

Theorem 5. *Any weighted graph of size n contains a spanning tree with scaling distortion $O(\sqrt{1/\epsilon})$. In particular, its ℓ_q-distortion is $O(1)$ for $1 \le q < 2$, $O(\sqrt{\log n})$ for $q = 2$, and $O(n^{1-2/q})$ for $2 < q \le \infty$.*

Given a graph, the spanning tree is created by recursively partitioning the metric space using a *hierarchical star partition*. The algorithm has three components, with the following high level description:

1. A decomposition algorithm that creates a single cluster. The decomposition algorithm is similar in spirit to the decomposition algorithm used in the previous section for metric spaces. We will later explain the main differences.

2. A star partition algorithm. This algorithm partitions a graph X into a central ball X_0 with center x_0 and a set of cones $X_1, \ldots, X_m$ and also outputs a set of edges of the graph $(y_1, x_1), \ldots, (y_m, x_m)$ that connect each cone set, $x_i \in X_i$ to the central ball, $y_i \in X_0$. The central ball is created by invoking the decomposition algorithm with a center x to obtain a cluster whose radius is in the range $[(1/2)\mathrm{rad}_{x_0}(X) \ldots (5/8)\mathrm{rad}_{x_0}(X)]$. Each cone set X_i is created by invoking the decomposition algorithm on the "cone-metric" obtained from x_0, x_i. Informally, a ball in the cone-metric around x_i with radius r is the set of all points x such that $d(x_0, x_i) + d(x_i, x) - d(x_0, x) \leq r$. Hence each cone X_i is a ball whose center is x_i in some appropriately defined "cone-metric". The radius of each ball in the cone metric is chosen to be $\approx \tau^k \mathrm{rad}_{x_0}(X)$ where $\tau < 1$ is some fixed constant and k is the depth of the recursion. Unfortunately, at some stage the radius may be too small for the decompose algorithm to preform well enough. In such cases we must reset the parameters that govern the radius of the cones. (in the next bullet, we will define more accurately how the recursion is performed and when this parameter of a cluster may be reset). The main property of this star decomposition is that for any point $x \in X_i$, the distance to the center x_0 does not increase by too much. More formally, $d_{X_0 \cup \{(y_i, x_i)\} \cup X_i}(x_0, x)/d(x_0, x) \leq \prod_{j \leq k}(1 + \tau^j)$ where k is the depth of the recursion. Informally, this property is used in order to obtain a constant blowup in the diameter of each cluster in the final spanning tree.

3. Recursive application of the star partition. As mentioned in the previous bullet, the radius of the balls in the cone metric are exponentially decreasing. However at certain stages in the recursion, the cone radius becomes too small and the parameters governing the cone radius must be reset. Clusters in which the parameters need to be restarted are called *reset clusters*. The two parameters that are associated with a reset cluster X are $n = |X|$, and $\Lambda = \mathrm{rad}(X)$. Specifically, a cluster is called a reset cluster if its size relative to the size of the last reset cluster is larger than some constant times its radius relative to radius of the last reset cluster. In that case n and Λ are updated to the values of the current cluster. This implies that reset clusters have small diameter, hence their total contribution to the increase of radius is small. Moreover, resetting the parameters allows the decompose algorithm to continue to produce the clusters with the necessary properties to obtain the desired scaling distortion. Using resets, the algorithm can continue recursively in this fashion until the spanning tree is formed.

Decompose algorithm. The decompose algorithm receives as input several parameters. First it obtains a pseudo-metric space (W, d) and point u (for the central ball this is just the shortest-paths metric, while for cones, this pseudo metric is the so called "cone-metric" which will be formally defined in the sequel). The goal of the decompose algorithm is to partition W into a cluster which is a ball $Z = B(u, r)$ and $\bar{Z} = W \setminus Z$.

Informally, this partition P is carefully chosen to maintain the scaling property: for every ϵ, the number of pairs whose distortion is too large is "small enough". Let $\hat{\Lambda}$ be a parameter corresponding to the radius of the cluster over which the star-partition is performed. Pairs that are separated by the partition may risk the possibility of being at distance $\Theta(\hat{\Lambda})$ in the constructed spanning tree. We denote by $\hat{B}_\epsilon(P)$ the number of pairs that may be distorted by at least $\Omega(\sqrt{1/\epsilon})$ if the distance between them will grow to $\hat{\Lambda}$. There are several parameters that control the number of pairs in $\hat{B}_\epsilon(P)$. Given a parameter $n \geq |W|$ which corresponds to the size of the last reset cluster containing W, we expect the number of "bad" pairs for a specific value of ϵ to be at most $O(\epsilon|Z| \cdot (n - |Z|))$. To allow to control this bound even tighter we have an additional parameter β so that the partition P will have the property that $\hat{B}_\epsilon(P) = O(\epsilon|Z| \cdot (n - |Z|) \cdot \beta)$. However, if we insist that this property holds true for all ϵ we cannot maintain a small enough bound on the maximum value for the radius r. Since this value determines the amount of increase in the radius of the cluster, we would like to be able to bound it. Therefore, we keep another parameter, denoted ϵ_{lim}. That is, the partition P will be good only for those values of ϵ satisfying $\epsilon \leq \epsilon_{\mathrm{lim}}$.

The radius r of the ball is controlled by the parameters $\hat{\Lambda}$, θ and a value $\alpha \leq \sqrt{\epsilon_{\mathrm{lim}}}$. The guarantee is

that $r \in [\theta\hat{\Lambda}, (\theta+\alpha)\hat{\Lambda}]$. Recall that $\hat{\Lambda}$, corresponds to the radius of the cluster over which the star-partition is performed. For the central ball of the star-partition θ is fixed to $1/2$ and for the star's cones θ is fixed to 0. Indeed, as indicated above, the value of $\epsilon_{\lim}$ determines the increase in the radius of the cluster by setting the value for α. This cannot, however, be set arbitrarily small, in order to satisfy all of the partition's properties, and so $\epsilon_{\lim}$ must be set above some minimum value of $|W|/(n \cdot \beta)$. Intuitively, we can only keep α small if $|W| \ll n$.

Let us explain now how the decompose algorithm will be used within our overall scheme. The parameter β is chosen such that it is bounded by μ^k where $\mu < 1$ is some fixed constant and k is the depth of the recursion from the last reset cluster. Hence, for every ϵ that is smaller than $\epsilon_{\lim}$, the property obtained by the decompose algorithm is that the number of newly distorted edges is at most $O(\epsilon|Z| \cdot (n-|Z|) \cdot \mu^k)$. For ϵ that are larger than $\epsilon_{\lim}$, we show that the number of points in the current cluster is less than an ϵ fraction of the number of points in the last reset cluster, hence we can discard all the pairs in such clusters and the total sum of all such discarded pairs is small. Therefore, the total number of distorted edges is bounded by summing the distorted edges over all clusters, for each cluster depending on whether ϵ is smaller or larger than $\epsilon_{\lim}$ of that cluster. The bound obtained also uses the fact that μ^k is a geometric series.

Now, if X is not a reset cluster then $|X|/n$ is small compared to the ratio of its radius and the radius of the last reset cluster. We show that this ratio drops exponentially, bounded by $(\frac{5}{8})^k$, where k is the depth of the recursion since the last reset cluster. By letting $\epsilon_{\lim} = |X|/(n \cdot \beta)$, and as $\mu < \frac{5}{8}$, we maintain that $\alpha \le \sqrt{\epsilon_{\lim}} = \tau^k$ for some $\tau < 1$, as we desired.

We now turn to the formal description of the algorithm and its analysis. We will make use of the following predefined constants: $c = 2e$, $c' = e(2e+1)$, $\hat{c} = 22$, and $C = 8\sqrt{c \cdot \hat{c}}$. Finally, the distortion is given by $\hat{C} = 150C \cdot c'$. For any $0 < \epsilon \le 1$ denote by $B_\epsilon(X)$ the total number of pairs $(x, y) \in X$ such that $d_T(x, y) > (\hat{C}/\sqrt{\epsilon})d_X(x, y)$. The exact properties of the decomposition algorithm is captured by the following Lemma:

Lemma 5. *Given a metric space (W, d), a point $u \in W$ and parameters $n \in \mathbb{N}$, $\hat{\Lambda} > 0$, and $\beta, \theta > 0$, there exists an algorithm $\texttt{decompose}((W, d), u, \hat{\Lambda}, \theta, n, \epsilon_{\lim}, \beta)$ that computes a partition $P = (Z; \bar{Z})$ of W such that $Z = B_{(W,d)}(u, r)$ and $r/\hat{\Lambda} \in [\theta, \theta+\alpha]$ where $\alpha = \sqrt{\epsilon_{\lim}}/C$. Let $\hat{B}_\epsilon(P) = |\{(x, y) \mid x \in Z \wedge y \in \bar{Z} \wedge d(x, y) \le \frac{\sqrt{\epsilon}\cdot\hat{\Lambda}}{150C}\}|$. For $n \ge |W|$ and $\epsilon_{\lim} \ge \frac{|W|}{\beta \cdot n}$ the partition has the property that for any $\epsilon \in (0, \epsilon_{\lim}]$:*

$$\hat{B}_\epsilon(P) \le \epsilon|Z| \cdot (n-|Z|) \cdot \beta.$$

Star-Partition algorithm. Consider a cluster X with center x_0 and parameters n, Λ. Recall that parameters n, Λ are the number of points and the radius (respectively) of the last reset cluster. A star-partition, partitions X into a central ball X_0, and cone-sets $X_1, \ldots, X_m$ and edges $(y_1, x_1), \ldots, (y_m, x_m)$, the value m is determined by the star-partition algorithm when no more cones are required. Each cone-set X_i is connected to X_0 by the edge (y_i, x_i), $y_i \in X_0, x_i \in X_i$. Denote by P_0 the partition creating the central ball X_0 and by $\{P_i\}_{i=1}^m$ the partitions creating the cones. In order to create the cone-set X_i use the decompose algorithm on the cone-metric $\ell_{x_i}^{x_0}$ defined below.

Definition 4 (cone metric[3]). *Given a metric space (X, d) set $Y \subset X$, $x \in X$, $y \in Y$ define the* cone-metric *$\ell_y^x : Y^2 \to \mathbb{R}^+$ as $\ell_y^x(u, v) = |(d(x, u) - d(y, u)) - (d(x, v) - d(y, v))|$.*

Note that $B_{(Y,\ell_y^x)}(y, r) = \{v \in Y | d(x, y) + d(y, v) - d(x, v) \le r\}$.

Hierarchical-Star-Partition algorithm. Given a graph $G = (X, E, \omega)$, create the tree by choosing some $x \in X$, setting X as a reset cluster and calling: $\texttt{hierarchical-star-partition}(X, x, |X|, \mathrm{rad}_x(X))$.

[3] In fact, the cone-metric is a pseudo-metric.

$(X_0, \ldots, X_m, (y_1, x_1), \ldots, (y_m, x_m)) =$ `star-partition`(X, x_0, n, Λ):

1. Set $i = 0$; $\beta = \frac{1}{\hat{c}}\left(\frac{\mathrm{rad}_{x_0}(X)}{\Lambda}\right)^{1/4}$; $\epsilon_{\lim} = |X|/(\beta n)$; $\hat{\Lambda} = \mathrm{rad}_{x_0}(X)$;
2. $(X_i, Y_i) =$ `decompose`$((X, d), x_0, \hat{\Lambda}, 1/2, \epsilon_{\lim}, \beta)$;
3. If $Y_i = \emptyset$ set $m = i$ and stop; Otherwise, set $i = i + 1$;
4. Let (x_i, y_i) be an edge in E such that $y_i \in X_0, x_i \in Y_{i-1}$;
5. Let $\ell = \ell_{x_i}^{x_0}$ be cone-metric of x_0, x_i on the subspace Y_{i-1};
6. $(X_i, Y_i) =$ `decompose`$((Y_{i-1}, \ell), x_i, \hat{\Lambda}, 0, \epsilon_{\lim}, \beta)$;
7. goto 3;

Figure 1: `star-partition` algorithm

$T =$ `hierarchical-star-partition`(X, x, n, Λ):

1. If $|X| = 1$ set $T = X$ and stop.
2. $(X_0, \ldots, X_m, (y_1, x_1), \ldots, (y_m, x_m)) =$ `star-partition`(X, x, n, Λ);
3. For each $i \in [1, \ldots, m]$:
4. If $\frac{|X_i|}{n} \leq c\frac{\mathrm{rad}_{x_i}(X_i)}{\Lambda}$ then $T_i =$ `hierarchical-star-partition`(X_i, x_i, n, Λ);
5. Otherwise, set X_i to be a **reset cluster**, $T_i =$ `hierarchical-star-partition`$(X_i, x_i, |X_i|, \mathrm{rad}_{x_i}(X_i))$;
6. Let T be the tree formed by connecting T_0 with T_i using edge (y_i, x_i) for each $i \in [1, \ldots, m]$;

Figure 2: `hierarchical-star-partition` algorithm

4.1 Algorithm Analysis

The hierarchical star-partition of $G = (X, E, \omega)$ naturally induces a laminar family $\mathcal{F} \subseteq 2^X$. Let $\mathcal{G}$ be the rooted *construction tree* whose nodes are sets in $\mathcal{F}$, $F \in \mathcal{F}$ is a parent of $F' \in \mathcal{F}$ if F' is a cluster formed by the partition of F. Observe that the spanning tree T obtained by our hierarchical star decomposition has the property that every $F \in \mathcal{F}$ corresponds to a sub tree $T[F]$ of T. Let $\mathcal{R} \subseteq \mathcal{F}$ be the set of all reset clusters. For each $F \in \mathcal{F}$, let $\mathcal{G}_F$ be the sub-tree of the construction tree $\mathcal{G}$ rooted at F, that contains all the nodes X whose path to F (excluding F and X) contains no node in $\mathcal{R}$. For $F \in \mathcal{F}$ let $\mathcal{R}(F) \subseteq \mathcal{R}$ be the set of reset cluster which are descendants of F in $\mathcal{G}_F$ (These are the leaves of the construction sub-tree $\mathcal{G}_F$ rooted at F). In what follows we use the following convention on our notation: whenever X is a cluster in $\mathcal{G}$ with center point x_0 with respect to which the star-partition of X has been constructed, we define $\mathrm{rad}(X) = \mathrm{rad}_{x_0}(X)$. We first claim the following bound on α produced by the decompose algorithms.

Claim 6. *Fix $F \in \mathcal{F}$ and $\mathcal{G}_F$. Let $X \in \mathcal{G}_F \setminus \mathcal{R}(F)$, such that $d_{\mathcal{G}}(X, F) = k$. By our construction, in each iteration of the partition algorithm the radius decreases by a factor of at least $\frac{5}{8}$, hence* $\mathrm{rad}(X) \leq \mathrm{rad}(F) \cdot (\frac{5}{8})^k$.

Proof. For any cluster F, the radius of the central ball in the star decomposition of F is at most $((1/2) + \alpha)\mathrm{rad}(F)$. Since the radius of this ball is also at least $(1/2)\mathrm{rad}(F)$ then the radius of each cone is at most $((1/2) + \alpha)\mathrm{rad}(F)$ as well. Let $Y \in \mathcal{R}$ such that $X \in \mathcal{G}_Y$. Since $C = 8\sqrt{c \cdot \hat{c}}$ then $\alpha = \sqrt{\epsilon_{\lim}}/C = \sqrt{\frac{|X|}{c|Y|}\left(\frac{\mathrm{rad}(Y)}{\mathrm{rad}(X)}\right)^{1/4}}/8 \leq \frac{1}{8}\sqrt{\left(\frac{\mathrm{rad}(X)}{\mathrm{rad}(Y)}\right)^{3/4}} \leq \frac{1}{8}$. $\square$

We now show that the spanning tree of each cluster increases its diameter by at most a constant factor. Recall that $c' = e(2e+1)$.

Lemma 7. *For every* $F \in \mathcal{F}$ *and* $T[F] \subseteq T$ *we have* $\mathrm{rad}(T[F]) \le c' \cdot \mathrm{rad}(F))$.

Proof. Let $Y \in \mathcal{R}$. We first prove by induction on the construction tree $\mathcal{G}$ that for every $X \in \mathcal{G}_Y$ with $t = d_{\mathcal{G}}(X,Y)$ we have

$$\mathrm{rad}(T[X]) \le \prod_{j\ge t}(1+\frac{1}{8}(\frac{7}{8})^j)\left(\mathrm{rad}(X) + \sum_{R\in\mathcal{R}(Y)\cap\mathcal{G}_X}\mathrm{rad}(T[R])\right) \tag{1}$$

Fix some cluster $X \in \mathcal{G}_Y$, such that $t = d_{\mathcal{G}}(X,Y)$ and assume the hypothesis is true for all its children in $\mathcal{G}_Y$. If X is a leaf of $\mathcal{G}_Y$ then it is a reset cluster and the claim trivially holds (since $X \in \mathcal{R}(Y)\cap\mathcal{G}_X$). Otherwise, assume we partition X into $X_0,\dots,X_m$. Let $i \in [1,m]$ such that X_i is the cluster such that $\omega(y_i,x_i) + \mathrm{rad}(T[X_i])$ is maximal, hence $\mathrm{rad}(T[X]) \le \mathrm{rad}(T[X_0]) + \omega(y_i,x_i) + \mathrm{rad}(T[X_i])$. There are four cases to consider depending on whether X_0 and X_i belong to $\mathcal{R}$. Here we show the case of $X_0, X_i \notin \mathcal{R}$, the other cases are similar and easier. Using Claim 6 we obtain the following bound on the increase in radius: $\alpha \le 1/8\sqrt{\left(\frac{\mathrm{rad}(X)}{\mathrm{rad}(Y)}\right)^{3/4}} \le 1/8(5/8)^{3t/8} \le 1/8(7/8)^t$. It follows that $\mathrm{rad}(X_0) + \omega(y_i,x_i) + \mathrm{rad}(X_i) \le \mathrm{rad}(X)(1+\alpha) \le \mathrm{rad}(X)(1+1/8(7/8)^t)$. By the induction hypothesis we know that $\mathrm{rad}(T[X_0]) \le \prod_{j\ge t+1}(1+\frac{1}{8}(\frac{7}{8})^j)(\mathrm{rad}(X_0) + \sum_{R\in\mathcal{R}(Y)\cap\mathcal{G}_{X_0}}\mathrm{rad}(T[R]))$ and $\mathrm{rad}(T[X_i]) \le \prod_{j\ge t+1}(1+\frac{1}{8}(\frac{7}{8})^j)(\mathrm{rad}(X_i) + \sum_{R\in\mathcal{R}(Y)\cap\mathcal{G}_{X_i}}\mathrm{rad}(T[R]))$, hence

$$\begin{aligned}
\mathrm{rad}(T[X]) &\le \mathrm{rad}(T[X_0]) + \omega(y_i,x_i) + \mathrm{rad}(T[X_i]) \\
&\le \prod_{j\ge t+1}(1+\frac{1}{8}(\frac{7}{8})^j)\left(\mathrm{rad}(X_0) + \omega(y_i,x_i) + \mathrm{rad}(X_i) + \sum_{R\in\mathcal{R}(Y)\cap\mathcal{G}_X}\mathrm{rad}(T[R])\right) \\
&\le \prod_{j\ge t+1}(1+\frac{1}{8}(\frac{7}{8})^j)\left(\mathrm{rad}(X)(1+\frac{1}{8}(\frac{7}{8})^t) + \sum_{R\in\mathcal{R}(Y)\cap\mathcal{G}_X}\mathrm{rad}(T[R])\right) \\
&\le \prod_{j\ge t}(1+\frac{1}{8}(\frac{7}{8})^j)\left(\mathrm{rad}(X) + \sum_{R\in\mathcal{R}(Y)\cap\mathcal{G}_X}\mathrm{rad}(T[R])\right).
\end{aligned}$$

This completes the proof of (1). Now we continue to prove the Lemma. First, we prove by induction on the construction tree $\mathcal{G}$ that the Lemma holds for the set of reset clusters. In fact we show a somewhat stronger bound. Recall that $c = 2e$. We show that for every cluster $Y \in \mathcal{R}$ we have $\mathrm{rad}(T[Y]) \le c\cdot\mathrm{rad}(Y)$. Assume the induction hypothesis is true for all descendants of Y in $\mathcal{R}$. In particular, for all $R \in \mathcal{R}(Y)$, $\mathrm{rad}(T[R]) \le c\cdot \mathrm{rad}(R)$. Recall that R becomes a reset cluster since $\mathrm{rad}(R) \le \frac{\mathrm{rad}(Y)}{c\cdot|Y|}|R|$, hence $\sum_{R\in\mathcal{R}(Y)}\mathrm{rad}(R) \le \mathrm{rad}(Y)/c$. Using (1) we have that

$$\begin{aligned}
\mathrm{rad}(T[Y]) &\le \prod_{j\ge 0}(1+\frac{1}{8}(\frac{7}{8})^j)\left(\mathrm{rad}(Y) + \sum_{R\in\mathcal{R}(Y)}\mathrm{rad}(T[R])\right) \\
&\le (e^{\frac{1}{8}\sum_{j\ge 0}(\frac{7}{8})^j})(\mathrm{rad}(Y) + c\cdot\mathrm{rad}(Y)/c) \\
&\le e\cdot 2\mathrm{rad}(Y) = c\cdot\mathrm{rad}(Y).
\end{aligned}$$

Finally, we show the Lemma holds for all the other clusters. Let $F \in \mathcal{F}\setminus\mathcal{R}$ and $Y \in \mathcal{R}$ such that $F \in \mathcal{G}_Y$. Let $t = d_{\mathcal{G}}(F,Y)$. Note that $\sum_{R\in\mathcal{R}(Y)\cap\mathcal{G}_F}|R| = |F|$. Since $F \notin \mathcal{R}$ we have $\frac{\mathrm{rad}(Y)}{c|Y|} \le \frac{\mathrm{rad}(F)}{|F|}$ hence

$$\sum_{R\in\mathcal{R}(Y)\cap\mathcal{G}_F}\mathrm{rad}(R) \le \frac{\mathrm{rad}(Y)}{c|Y|}\sum_{R\in\mathcal{R}(Y)\cap\mathcal{G}_F}|R| \le \mathrm{rad}(F).$$

By (1) and the second induction we get

$$\begin{aligned}
\mathrm{rad}(T[F]) &\le \prod_{j\ge t}\left(1+\frac{1}{8}\left(\frac{7}{8}\right)^j\right)\left(\mathrm{rad}(F)+\sum_{R\in\mathcal{R}(Y)\cap\mathcal{G}_F}\mathrm{rad}(T[R])\right)\\
&\le e\cdot\left(\mathrm{rad}(F)+c\sum_{R\in\mathcal{R}(Y)\cap\mathcal{G}_F}\mathrm{rad}(R)\right)\\
&\le e\cdot\mathrm{rad}(F)(c+1)=c'\cdot\mathrm{rad}(F),
\end{aligned}$$

proving the Lemma. ☐

We now proceed to bound for every ϵ the number of pairs with distortion $\Omega(\sqrt{1/\epsilon})$, thus proving the scaling distortion of our constructed the spanning tree. We begin with some definitions that will be crucial in the analysis.

Definition 5. *For each $\epsilon\in(0,1]$ and $R\in\mathcal{R}$ let $\mathcal{K}(R,\epsilon)=\{F\in\mathcal{G}_R \mid |F|<\epsilon/\hat{c}\cdot|R|\}$.*

Hence, a cluster is in $\mathcal{K}(R,\epsilon)$ if it contains less than $\epsilon/\hat{c}$ fraction of the points of R. Informally, when counting the badly distorted edges for a given ϵ, whenever we reach a cluster in $\mathcal{K}(R,\epsilon)$ we count all its pairs as bad. If $X\in\mathcal{G}_R$ then let $\mathcal{K}(X,\epsilon)=\mathcal{K}(R,\epsilon)\cap\mathcal{G}_X$. For $R\in\mathcal{R}$ let $\mathcal{G}_{R,\epsilon}$ be the sub-tree rooted at R, that contains all the nodes X whose path to R (excluding R and X) contains no node in $\mathcal{R}\cup\mathcal{K}(R,\epsilon)$. Observe that $\mathcal{G}_{R,\epsilon}$ is a sub tree of $\mathcal{G}_R$.

Lemma 8. *For any $R\in\mathcal{R}$, $\epsilon\in(0,1]$ we have that $B_\epsilon(R)\le\epsilon\binom{R}{2}$.*

Proof. Fix some $\epsilon\in(0,1]$. Fix $F\in\mathcal{R}$. In order to prove the claim for F, we will first prove the following inductive claim for all $X\in\mathcal{G}_F$. Let $t=d_{\mathcal{G}}(X,F)$. Let $\mathcal{E}(X)=\left(\binom{X}{2}\setminus\bigcup_{R\in\mathcal{R}(X)}\binom{R}{2}\cup\bigcup_{K\in\mathcal{K}(X,\epsilon)}\binom{K}{2}\right)$.

$$B_\epsilon(X)\le\frac{2}{\hat{c}}\cdot\epsilon\sum_{i\ge t}(9/10)^i\cdot|\mathcal{E}(X)|+\sum_{R\in\mathcal{R}(F)\cap\mathcal{G}_X}B_\epsilon(R)+\sum_{K\in\mathcal{K}(F,\epsilon)\cap\mathcal{G}_X}B_\epsilon(K). \tag{2}$$

The base of the induction, where X is a leaf in $\mathcal{G}_F$, *i.e.* $X\in\mathcal{R}(F)\cup\mathcal{K}(F,\epsilon)$, is trivial. Assume the claim holds for all the children $X_0,\ldots,X_m$ of X. Let $P=\{P_i\}_{i=0}^m$ be the star-partition of X, where $P_i=(X_i,Y_i)$, $Y_i=\cup_{j=i+1}^m X_j$. Recall the definition of $\hat{B}_\epsilon(P_i)=|\{(x,y)\mid x\in X_i\wedge y\in Y_i\wedge d(x,y)\le\frac{\sqrt{\epsilon}\cdot\hat{\Lambda}}{150C}\}|$, where $\hat{\Lambda}=\mathrm{rad}(X)$. Denote $\hat{B}_\epsilon(P)=\sum_{i=0}^m\hat{B}_\epsilon(P_i)$. By Lemma 7 we have that $\mathrm{rad}(T(X))\le c'\mathrm{rad}(X)$. Hence, the number of pairs distorted more than $150C\cdot c'\sqrt{1/\epsilon}$ by the partition P is bounded by $\hat{B}_\epsilon(P)$. Now, since $X\notin\mathcal{K}(F,\epsilon)$ then $\epsilon<\hat{c}\cdot|X|/|F|\le 1/\beta\cdot|X|/|F|=\epsilon_{\lim}$. Hence we can apply Lemma 5 to deduce a bound on $B_\epsilon(P_i)$. By Claim 6 we have $\beta=\frac{1}{\hat{c}}\left(\frac{\mathrm{rad}(X)}{\mathrm{rad}(F)}\right)^{1/4}\le\frac{1}{\hat{c}}(\frac{5}{8})^{t/4}$. From Lemma 5 we obtain

$$\hat{B}_\epsilon(P)=\sum_{i=0}^m\hat{B}_\epsilon(P_i)\le\frac{1}{\hat{c}}\cdot\epsilon\left(\frac{5}{8}\right)^{t/4}\sum_{i=0}^m|X_i||F\setminus X_i|\le\frac{2}{\hat{c}}\cdot\epsilon(9/10)^t|\mathcal{E}(X)|.$$

Using the induction hypothesis we get that

$$\begin{aligned}
B_\epsilon(X) &\leq \hat{B}_\epsilon(P) + \sum_{j=0}^{m} B_\epsilon(X_j) \\
&\leq \frac{2}{\hat{c}} \cdot \epsilon (9/10)^t |\mathcal{E}(X)| + \sum_{j=0}^{m} \left(\frac{2}{\hat{c}} \cdot \epsilon |\mathcal{E}(X_j)| \sum_{i \geq t+1} (9/10)^i + \sum_{R \in \mathcal{R}(F) \cap \mathcal{G}_{X_j}} B_\epsilon(R) + \sum_{K \in \mathcal{K}(F,\epsilon) \cap \mathcal{G}_{X_j}} B_\epsilon(K) \right) \\
&\leq \frac{2}{\hat{c}} \cdot \epsilon (9/10)^t |\mathcal{E}(X)| + \frac{2}{\hat{c}} \cdot \epsilon |\mathcal{E}(X)| \sum_{i \geq t+1} (9/10)^i + \sum_{R \in \mathcal{R}(F) \cap \mathcal{G}_X} B_\epsilon(R) + \sum_{K \in \mathcal{K}(F,\epsilon) \cap \mathcal{G}_X} B_\epsilon(K) \\
&\leq \frac{2}{\hat{c}} \cdot \epsilon \sum_{i \geq t} (9/10)^i |\mathcal{E}(X)| + \sum_{R \in \mathcal{R}(F) \cap \mathcal{G}_X} B_\epsilon(R) + \sum_{K \in \mathcal{K}(F,\epsilon) \cap \mathcal{G}_X} B_\epsilon(K),
\end{aligned}$$

which proves the inductive claim. We now prove the Lemma by induction on the construction tree $\mathcal{G}$. Let $F \in \mathcal{R}$. By the induction hypothesis $B_\epsilon(R) \leq \epsilon \binom{R}{2}$ for every $R \in \mathcal{R}(F)$. Observe that if $K \in \mathcal{K}(F, \epsilon)$ then we discard all pairs in K. Hence $B_\epsilon(K) \leq |K|^2 \leq \frac{1}{\hat{c}} \cdot \epsilon |F| \cdot |K|$. Recall that $\hat{c} = 22$. From (2) we obtain

$$\begin{aligned}
B_\epsilon(F) &\leq \frac{2}{22} \cdot \epsilon \sum_{i \geq 0} (9/10)^i \cdot |\mathcal{E}(F)| + \epsilon \sum_{R \in \mathcal{R}(F)} \binom{R}{2} + \sum_{K \in \mathcal{K}(F,\epsilon)} \frac{1}{22} \cdot \epsilon |F| \cdot |K| \\
&\leq \left[\frac{20}{22} \cdot \epsilon \cdot |\mathcal{E}(F)| + \frac{20}{22} \epsilon \sum_{R \in \mathcal{R}(F)} \binom{R}{2} \right] + \left[\frac{2}{22} \epsilon \sum_{R \in \mathcal{R}(F)} |R| \cdot (|R|-1)/2 + \frac{1}{22} \epsilon \cdot |F| \sum_{K \in \mathcal{K}(F,\epsilon)} |K| \right] \\
&\leq \frac{20}{22} \epsilon \binom{|F|}{2} + \frac{1}{22} \epsilon \cdot |F| \left(\sum_{R \in \mathcal{R}(F)} (|R|-1) + \sum_{K \in \mathcal{K}(F,\epsilon)} |K| \right) \\
&\leq \frac{20}{22} \epsilon \binom{|F|}{2} + \frac{1}{22} \epsilon \cdot |F|(|F|-1) \\
&= \epsilon \binom{|F|}{2},
\end{aligned}$$

where the third inequality follows from the definition of $\mathcal{E}(X)$ and from the fact that for each $K \in \mathcal{K}(F, \epsilon)$, $R \in \mathcal{R}(F)$ we have $K \cap R = \emptyset$. ☐

Applying Lemma 8 on the original graph proves Theorem 2. Finally, we complete the proof of Lemma 5 stating the properties of our generic decompose algorithm.

Proof of Lemma 5. We distinguish between the following two cases:

Case 1: $|B(u, (\theta + \alpha/2)\hat{\Lambda})| \leq n/2$. In this case let $\hat{\epsilon} = \max\{\epsilon \in (0, \epsilon_{\lim}] \mid |B(u, (\theta + \frac{\sqrt{\epsilon}}{4C})\hat{\Lambda})| \geq \epsilon \cdot \beta \cdot n\}$. Let $\hat{S} = [(\theta + \frac{\sqrt{\epsilon}}{4C})\hat{\Lambda}, (\theta + \frac{\sqrt{\epsilon}}{2C})\hat{\Lambda})$, and $S = \left[\left(\theta + \frac{\sqrt{\epsilon}}{C}\left(\frac{1}{4} + \frac{1}{25}\right)\right)\hat{\Lambda}, \left(\theta + \frac{\sqrt{\epsilon}}{C}\left(\frac{1}{2} - \frac{1}{25}\right)\right)\hat{\Lambda}\right]$.

Case 2: $|B(u, (\theta+\alpha/2)\hat{\Lambda})| > n/2$. In this case let $\hat{\epsilon} = \max\{\epsilon \in [0, \epsilon_{\lim}] \mid |W \setminus B(u, (\theta+\alpha-\frac{\sqrt{\epsilon}}{4C})\hat{\Lambda})| \geq \epsilon \cdot \beta \cdot n\}$. Let $\hat{S} = [(\theta + \alpha - \frac{\sqrt{\epsilon}}{2C})\hat{\Lambda}, (\theta + \alpha - \frac{\sqrt{\epsilon}}{4C})\hat{\Lambda}]$, and $S = \left[\left(\theta+\alpha-\frac{\sqrt{\epsilon}}{C}\left(\frac{1}{2} - \frac{1}{25}\right)\right)\hat{\Lambda}, \left(\theta+\alpha-\frac{\sqrt{\epsilon}}{C}\left(\frac{1}{4} + \frac{1}{25}\right)\right)\hat{\Lambda}\right]$.

We show that one can choose $r \in S$ and define $Z = B(u, r)$ such that the property of the Lemma holds. We now show the property of the Lemma holds for all $\epsilon \in (32\hat{\epsilon}, \epsilon_{\lim}]$ and any $r \in S$.

Proof for Case 1: In this case we will use the bound:

$$\hat{B}_\epsilon(P) \leq |B(u, r + \sqrt{\epsilon}\hat{\Lambda}/(150C)) \setminus Z| \cdot |Z|. \tag{3}$$

Note that $r+\sqrt{\epsilon}\hat{\Lambda}/(150C) \leq (\theta+\sqrt{\hat{\epsilon}}/(2C))\hat{\Lambda}+\sqrt{\epsilon/2}\hat{\Lambda}/(8C) \leq (\theta+\sqrt{\epsilon/2}/(4C))\hat{\Lambda}$, using that $\hat{\epsilon} \leq \epsilon/32$. Now, by the maximality of $\hat{\epsilon}$ we have $|B(u,(\theta+\sqrt{\epsilon/2}/(4C))\hat{\Lambda})| \leq \epsilon/2 \cdot \beta \cdot n$. Therefore, using (3) we get

$$\begin{aligned} \hat{B}_\epsilon(P) &\leq |B(u,(\theta+\sqrt{\epsilon/2}/(4C)\hat{\Lambda})| \cdot |Z| \\ &\leq (\epsilon \cdot \beta \cdot n/2) \cdot |Z| \leq \epsilon \cdot \beta \cdot |Z| \cdot (n-|Z|), \end{aligned}$$

using $|Z| \leq n/2$.

Proof for Case 2: In this case we will use the bound:

$$\hat{B}_\epsilon(P) \leq |\bar{Z}| \cdot |W \setminus B(u, r-\sqrt{\epsilon}\hat{\Lambda}/(150C))|. \tag{4}$$

Note that $r-\sqrt{\epsilon}\hat{\Lambda}/(150C) \geq (\theta+\alpha-\sqrt{\hat{\epsilon}}/(2C))\hat{\Lambda}-\sqrt{\epsilon/2}\hat{\Lambda}/(8C) \geq (\theta+\alpha-\sqrt{\epsilon/2}\hat{\Lambda}/(4C))\hat{\Lambda}$, using that $\hat{\epsilon} \leq \epsilon/32$. Now, from the maximality of $\hat{\epsilon}$ we have $|W \setminus B(u,(\theta+\alpha-\sqrt{\epsilon/2}/(4C))\hat{\Lambda})| < \epsilon \cdot \beta \cdot n/2$. Therefore, using (4) we get

$$\begin{aligned} \hat{B}_\epsilon(P) &\leq |\bar{Z}| \cdot |W \setminus B(u,(\theta+\alpha-\sqrt{\epsilon/2}/(4C))\hat{\Lambda})| \\ &\leq |\bar{Z}| \cdot \epsilon \cdot \beta \cdot n/2 \leq \epsilon \cdot \beta \cdot |Z|(n-|Z|), \end{aligned}$$

using $|Z| \geq n/2$.

We next show the property of the Lemma hold for all $\epsilon \in (0, 32\hat{\epsilon}]$. We will prove the claim for Case 1. The argument for Case 2 is the analogous. As before we define $Q = \{w \mid d(u,w) \in \hat{S}\}$. Now we have

Claim 9. *$|Q| \leq 4 \cdot \hat{\epsilon} \cdot \beta \cdot n$.*

Proof. We have $Q \subseteq B(u,(\theta+\sqrt{\hat{\epsilon}}/(2C))\hat{\Lambda})$. We distinguish between 2 cases: If $\hat{\epsilon} \leq \epsilon_{\lim}/4$ then $|B(u,(\theta+\sqrt{4\hat{\epsilon}}/(4C))\hat{\Lambda})| \leq 4\hat{\epsilon} \cdot \beta \cdot n$ (by the maximality of $\hat{\epsilon}$). Otherwise, $\hat{\epsilon} \in (\epsilon_{\lim}/4, \epsilon_{\lim}]$. In this case $|Q| \leq |W| \leq \epsilon_{\lim} \cdot \beta \cdot n \leq 4\hat{\epsilon} \cdot \beta \cdot n$. ☐

As before we will choose some $r \in S$ and the partition P will be $Z = B(u,r)$, $\bar{Z} = W \setminus Z$. It is easy to check that for any $r \in S$ we get $\hat{\epsilon} \cdot n \cdot \beta \leq |Z| \leq n/2$. We now find $r \in S$ which satisfy the property of the Lemma for all $0 < \epsilon \leq 32\hat{\epsilon}$: For any $r \in S$ and $\epsilon \leq 32\hat{\epsilon}$ let $S_r(\epsilon) = [r-\sqrt{\epsilon}\hat{\Lambda}/(150C)), r+\sqrt{\epsilon}\hat{\Lambda}/(150C))]$, $s(\epsilon) = \sqrt{\epsilon}\hat{\Lambda}/(75C)$ and let $Q_r(\epsilon) = \{w \mid d(u,w) \in S_r(\epsilon)\}$. Note that the length of the interval S is given by $s = 17/(100C)\sqrt{\hat{\epsilon}}\hat{\Lambda}$. We say that properly $A_r(\epsilon)$ holds if cutting at radius r is "good" for ϵ, formally: $A_r(\epsilon)$ iff $|Q_r(\epsilon)| \leq \sqrt{\epsilon \cdot \hat{\epsilon}/2} \cdot n \cdot \beta$. Notice that only pairs (x,y) such that $x, y \in Q_r(\epsilon)$ may be distorted by more than $150C\sqrt{1/\epsilon}$.

Claim 10. *There exists some $r \in S$ such that properly $A_r(\epsilon)$ holds for all $\epsilon \in (0, 32\hat{\epsilon}]$.*

Proof. As the proof of Claim 3 goes, we conduct exactly the same iterative process that greedily deletes the "worst" interval in S, which are $\{S_{r_j}(\epsilon_j)\}_{j=1}^t$, and we remain with $I_t \subseteq S$. We now argue that $I_t \neq \emptyset$. As before we have $\sum_{j=1}^t |Q_{r_j}(\epsilon_j)| \leq 2|Q| \leq 8\hat{\epsilon} \cdot \beta \cdot n$. Recall that since $A_{r_j}(\epsilon_j)$ does not hold then for any $1 \leq j \leq t$: $|Q_{r_j}(\epsilon_j)| > \sqrt{\epsilon_j \cdot \hat{\epsilon}/2} \cdot \beta \cdot n$ which implies that $\sum_{j=1}^t \sqrt{\epsilon_j} < 12\sqrt{\hat{\epsilon}}$. On the other hand, by definition

$$\sum_{j=1}^t s(\epsilon_j) \leq \sum_{j=1}^t \sqrt{\epsilon_j}\Delta/(75C) \leq 12/(75C) \cdot \sqrt{\hat{\epsilon}}\Delta = 16/(100C) \cdot \sqrt{\hat{\epsilon}}\Delta.$$

Since $s = 17/(100C) \cdot \sqrt{\hat{\epsilon}}\Delta$ then indeed $I_t \neq \emptyset$ so any $r \in I_t$ satisfies the condition of the claim. ☐

Claim 10 shows that for any $\epsilon \in (0, 32\hat{\epsilon}]$ we have

$$\hat{B}_\epsilon(P) \leq \epsilon \cdot \hat{\epsilon}/2 \cdot (n \cdot \beta)^2 \leq \epsilon \cdot \beta \cdot |Z| \cdot (n-|Z|),$$

which concludes the proof of the lemma. ☐

5 Probabilistic Scaling Embedding into spanning trees

The proof of this theorem is based on a somewhat simpler variation of the decomposition algorithm from the previous section. In fact, the `hierarchical-star-partition` algorithm remains practically the same, with modified sub-method `probabilistic-star-partition` given in Figure 3, instead of `star-partition`.

Let $f : \mathbb{R} \to \mathbb{R}_+$ be a monotone non-decreasing function satisfying

$$\int_1^\infty \frac{dx}{f(x)} = 1 \ . \tag{5}$$

For example if we define $\log^{(0)} n = n$, and for any $i > 0$ define recursively $\log^{(i)} n = \log(\log^{(i-1)} n)$, then we can take for any constants $\theta > 0$, $t \in \mathbb{N}$ the function $f(n) = \hat{c} \prod_{j=0}^{t-1} \log^{(j)}(n) \cdot \left(\log^{(t)}(n)\right)^{1+\theta}$, for sufficiently small constant $\hat{c} > 0$, and it will satisfy the conditions.

$(X_0, \ldots, X_t, (y_1, x_1), \ldots, (y_t, x_t)) =$ `probabilistic-star-partition`(X, x_0, Λ):

1. Set $k = 0$; $\hat{\Lambda} = \mathrm{rad}_{x_0}(X)$; $\alpha = \frac{1}{f(\log(2\Lambda/\hat{\Lambda}))}$;
2. Choose uniformly at random $\beta \in [0, 1/8]$.
3. Let γ be the value in $\{0, 1/16\}$ minimizing $|B(x_0, (1/2 + \gamma + 1/16)\hat{\Lambda})| - |B(x_0, (1/2 + \gamma)\hat{\Lambda})|$.
4. $X_0 = B(x_0, (1/2 + 3\gamma/2 + \beta/4)\hat{\Lambda})$; $Y_0 = X \setminus X_0$;
5. If $Y_k = \emptyset$ set $t = k$ and stop; Otherwise, set $k = k + 1$;
6. Let $v_k \in Y_{k-1}$ be the point minimizing $\hat{\chi}_k = \frac{|Y_0|}{|B_{Y_0}(x, \alpha\hat{\Lambda}/64)|}$; Set $\chi_k = \max\{4, \hat{\chi}_k\}$;
7. Choose $r \in [\alpha\hat{\Lambda}/16, \alpha\hat{\Lambda}/8]$ according to the distribution $p(r) = \frac{\chi_k^2}{1-\chi_k^{-2}} \frac{32 \ln \chi_k}{\alpha\hat{\Lambda}} \cdot \chi_k^{-32r/(\alpha\hat{\Lambda})}$;
8. Let (x_k, y_k) be the edge in E which lies on a shortest path from v_k to x_0 such that $y_k \in X_0, x_k \in Y_{k-1}$ [a];
9. Let $\ell = \ell_{x_k}^{x_0}$ be the cone-metric with respect to x_0 and x_k on the subspace Y_{k-1};
   $X_k = B_{(Y_{k-1}, \ell)}(x_k, r)$; $Y_k = Y_{k-1} \setminus X_k$.
10. goto 4;

[a] By the definition of cone-metric, if $z_k \in Y_{k-1}$ all the points on any shortest path from v_i to x_0 are either in X_0 or in Y_{k-1}

Figure 3: `probabilistic-star-partition` algorithm

5.1 Algorithm Analysis

Let $\hat{\mathcal{H}}$ be the distribution on laminar families induced by the algorithm above. Let $\mathcal{H} = supp(\hat{\mathcal{H}})$. We have the following analogs of Claim 6 and Lemma 7.

Claim 11. *Fix $\mathcal{F} \in \mathcal{H}$, $F \in \mathcal{F}$. Let $X \in \mathcal{G}_F \setminus \mathcal{R}(F)$, such that $d_{\mathcal{G}}(X, F) = k$. By our construction, in each iteration of the partition algorithm the radius decreases by a factor of at least $5/8$. Hence*

$$\mathrm{rad}(X) \le \mathrm{rad}(F) \cdot (5/8)^k \ .$$

Proof. For any cluster F, the radius of the central ball in the star decomposition of F is at most $(5/8)\mathrm{rad}(F)$. Since the radius of this ball is also at least $(1/2)\mathrm{rad}(F)$ then the radius of each cone is at most $((1/2) + \alpha/8)\mathrm{rad}(F) \le (5/8)\mathrm{rad}(F)$ as well. $\square$

We now show that the spanning tree of each cluster increases its diameter by at most a constant factor. Recall that $c' = e(2e+1)$.

Lemma 12. *For every $\mathcal{F} \in \mathcal{H}$, $F \in \mathcal{F}$ we have* $\mathrm{rad}(T[F]) \leq c' \cdot \mathrm{rad}(F)$.

Proof. Let $Y \in \mathcal{R}$. We first prove by induction on the construction tree $\mathcal{G}$ that for every $X \in \mathcal{G}_Y$ with $t = d_{\mathcal{G}}(X,Y)$ we have

$$\mathrm{rad}(T[X]) \leq \prod_{j \geq t}(1 + 1/(8f(1+j/5))) \left(\mathrm{rad}(X) + \sum_{R \in \mathcal{R}(Y) \cap \mathcal{G}_X} \mathrm{rad}(T[R]) \right) \tag{6}$$

Fix some cluster $X \in \mathcal{G}_Y$, such that $t = d_{\mathcal{G}}(X,Y)$ and assume the hypothesis is true for all its children in $\mathcal{G}_Y$. If X is a leaf of $\mathcal{G}_Y$ then it is a reset cluster and the claim trivially holds (since $X \in \mathcal{R}(Y) \cap \mathcal{G}_X$). Otherwise, assume we partition X into $X_0, \ldots, X_m$. Let $i \in [1,m]$ such that X_i is the cluster such that $\omega(y_i, x_i) + \mathrm{rad}(T[X_i])$ is maximal, hence $\mathrm{rad}(T[X]) \leq \mathrm{rad}(T[X_0]) + \omega(y_i,x_i) + \mathrm{rad}(T[X_i])$. There are four cases to consider depending on whether X_0 and X_i belong to $\mathcal{R}$. Here we show the case of $X_0, X_i \notin \mathcal{R}$, the other cases are similar and easier. Using Claim 11 $\log(2\mathrm{rad}(Y)/\mathrm{rad}(X)) \geq 1 + t/5$, it follows that

$$\mathrm{rad}(X_0) + \omega(y_i,x_i) + \mathrm{rad}(X_i) \leq \mathrm{rad}(X)\left(1 + 1/(8f(\log(2\mathrm{rad}(Y)/\mathrm{rad}(X))))\right) \leq \mathrm{rad}(X)\left(1 + 1/(8f(1+t/5))\right)$$

By the induction hypothesis we know that $\mathrm{rad}(T[X_0]) \leq \prod_{j \geq t+1}(1+1/(8f(1+j/5)))(\mathrm{rad}(X_0) + \sum_{R \in \mathcal{R}(Y) \cap \mathcal{G}_{X_0}} \mathrm{rad}(T[R]))$ and $\mathrm{rad}(T[X_i]) \leq \prod_{j \geq t+1}(1+1/(8f(1+j/5)))(\mathrm{rad}(X_i) + \sum_{R \in \mathcal{R}(Y) \cap \mathcal{G}_{X_i}} \mathrm{rad}(T[R]))$, hence

$$\begin{aligned}
\mathrm{rad}(T[X]) &\leq \mathrm{rad}(T[X_0]) + \omega(y_i,x_i) + \mathrm{rad}(T[X_i]) \\
&\leq \prod_{j \geq t+1}(1+1/(8f(1+j/5))) \left(\mathrm{rad}(X_0) + \omega(y_i,x_i) + \mathrm{rad}(X_i) + \sum_{R \in \mathcal{R}(Y) \cap \mathcal{G}_X} \mathrm{rad}(T[R]) \right) \\
&\leq \prod_{j \geq t+1}(1+1/(8f(1+j/5))) \left(\mathrm{rad}(X)(1+1/(8f(1+t/5))) + \sum_{R \in \mathcal{R}(Y) \cap \mathcal{G}_X} \mathrm{rad}(T[R]) \right) \\
&\leq \prod_{j \geq t}(1+1/(8f(1+j/5))) \left(\mathrm{rad}(X) + \sum_{R \in \mathcal{R}(Y) \cap \mathcal{G}_X} \mathrm{rad}(T[R]) \right).
\end{aligned}$$

This completes the proof of (6). Now we continue to prove the Lemma. First, we prove by induction on the construction tree $\mathcal{G}$ that the Lemma holds for the set of reset clusters. In fact we show a somewhat stronger bound. Recall that $c = 2e$. We show that for every cluster $Y \in \mathcal{R}$ we have $\mathrm{rad}(T[Y]) \leq c \cdot \mathrm{rad}(Y)$. Assume the induction hypothesis is true for all descendants of Y in $\mathcal{R}$. In particular, for all $R \in \mathcal{R}(Y)$, $\mathrm{rad}(T[R]) \leq c \cdot \mathrm{rad}(R)$. Recall that R becomes a reset cluster since $\mathrm{rad}(R) \leq \frac{\mathrm{rad}(Y)}{c \cdot |Y|}|R|$, hence $\sum_{R \in \mathcal{R}(Y)} \mathrm{rad}(R) \leq \mathrm{rad}(Y)/c$. Using Equation 6 and then Equation 5, we have that

$$\begin{aligned}
\mathrm{rad}(T[Y]) &\leq \prod_{j \geq 0}(1+1/(8f(1+j/5))) \left(\mathrm{rad}(Y) + \sum_{R \in \mathcal{R}(Y)} \mathrm{rad}(T[R]) \right) \\
&\leq (e^{1/8 \sum_{j \geq 0} 1/f(1+j/5)})(\mathrm{rad}(Y) + c \cdot \mathrm{rad}(Y)/c) \\
&\leq e^{5/8} \cdot 2\mathrm{rad}(Y) =\leq c \cdot \mathrm{rad}(Y).
\end{aligned}$$

Finally, we show the Lemma holds for all the other clusters. Let $F \in \mathcal{F} \setminus \mathcal{R}$ and $Y \in \mathcal{R}$ such that $F \in \mathcal{G}_Y$. Let $t = d_{\mathcal{G}}(F,Y)$. Note that $\sum_{R \in \mathcal{R}(Y) \cap \mathcal{G}_F} |R| = |F|$. Since $F \notin \mathcal{R}$ we have $\frac{\mathrm{rad}(Y)}{c|Y|} \leq \frac{\mathrm{rad}(F)}{|F|}$ hence

$$\sum_{R \in \mathcal{R}(Y) \cap \mathcal{G}_F} \mathrm{rad}(R) \leq \frac{\mathrm{rad}(Y)}{c|Y|} \sum_{R \in \mathcal{R}(Y) \cap \mathcal{G}_F} |R| \leq \mathrm{rad}(F).$$

By (6) and the second induction we get

$$\begin{aligned}
\operatorname{rad}(T[F]) &\le \prod_{j\ge t}(1+1/(8f(j/5)))\left(\operatorname{rad}(F)+\sum_{R\in\mathcal{R}(Y)\cap\mathcal{G}_F}\operatorname{rad}(T[R])\right)\\
&\le e\cdot\left(\operatorname{rad}(F)+c\sum_{R\in\mathcal{R}(Y)\cap\mathcal{G}_F}\operatorname{rad}(R)\right)\\
&\le e\cdot\operatorname{rad}(F)(c+1)=c'\cdot\operatorname{rad}(F),
\end{aligned}$$

proving the Lemma. □

For any $i>0$ let $\hat{\mathcal{H}}^{(i)}$ be the distribution on laminar families induced by i iterations of our probabilistic `hierarchical-star-partition` algorithm. Let $\mathcal{H}^{(i)} = supp(\hat{\mathcal{H}}^{(i)})$. Given $\mathcal{F}^{(i)}\in\mathcal{H}^{(i)}$. Let $\mathcal{G}^{(i)}$ be the corresponding construction tree of $\mathcal{F}^{(i)}$. Given $\mathcal{F}^{(i)}$, for any $x\in X$ let $F_i(x)$ be the leaf in $\mathcal{G}^{(i)}$ containing x.

Given $x,y\in G$ and $j>0$ define events $\mathcal{C},\mathcal{C}_{\text{ball}},\mathcal{X},\mathcal{Y},\mathcal{Z}$ as follows:

- Let $\mathcal{C}(x,y,j)$ be the event that there exists $i>0$ and $\mathcal{F}^{(i)}\in\mathcal{H}^{(i)}$ such that the following holds:
  1. $(\frac{8}{5})^j \le \operatorname{rad}(F_i(x)) < (8/5)^{j+1}$.
  2. $B(x,d(x,y))\subseteq F_{i+1}(x)$.
  3. $B(x,d(x,y))\nsubseteq F_i(x)$.
- Let $\mathcal{C}_{\text{ball}}(x,y,j)$ be the event that there exists $i>0$ and $\mathcal{F}^{(i)}\in\mathcal{H}^{(i)}$ such that the following holds:
  1. $(\frac{8}{5})^j \le \operatorname{rad}(F_i(x)) < (8/5)^{j+1}$.
  2. $X_0 = B_{F_i(x)}(x_0,r)$ and r chosen as in the algorithm.
  3. $B(x,d(x,y))\bowtie(X_0,F_i(x)\setminus X_0)$.
  4. $B(x,d(x,y))\subseteq F_{i+1}(x)$.

  Notice that by Claim 11 for each $\mathcal{F}$, the first property holds for at most one value of i, denote this value by $i_j(\mathcal{F})$.
- Let $\mathcal{X}(x,y,j,Z)$ be the event that there exist $i>0$ and $\mathcal{F}^{(i)}\in\mathcal{H}^{(i)}$ such that the following holds:
  1. $(\frac{8}{5})^j \le \operatorname{rad}(F_i(x)) < (8/5)^{j+1}$.
  2. $Z = B_{F_i(x)}(x_0,r)$ and r chosen as in the algorithm.
  3. $B(x,d(x,y))\subseteq Z$.
- Let $\mathcal{Y}(x,y,j,Z)$ be the event that there exist $i>0$ and $\mathcal{F}^{(i)}\in\mathcal{H}^{(i)}$ such that the following holds:
  1. $(\frac{8}{5})^j \le \operatorname{rad}(F_i(x)) < (8/5)^{j+1}$.
  2. $Z = F_i(x)\setminus B_{F_i(x)}(x_0,r)$ and r chosen as in the algorithm.
  3. $B(x,d(x,y))\subseteq Z$.
- Let $\mathcal{Z}(x,y,j,Z)=\mathcal{Y}(x,y,j,Z)\cup\mathcal{X}(x,y,j,Z)$.

We omit the parameters x,y,j (or part of them) from $\mathcal{C},\mathcal{C}_{\text{ball}},\mathcal{X},\mathcal{Y},\mathcal{Z}$ when clear from context. Here is an informal description of events $\mathcal{C},\mathcal{C}_{\text{ball}},\mathcal{X},\mathcal{Y},\mathcal{Z}$. Fix x,y and let $B=B(x,d(x,y))$. Event $\mathcal{C}(j)$ is the event that the first time that B is cut is when the parent cluster has radius $\approx(8/5)^j$. Event $\mathcal{C}_{\text{ball}}(j)$ is the event that the first time that B is cut is by the central ball given that that the parent cluster has radius $\approx(8/5)^j$, observe that $\mathcal{C}_{\text{ball}}(j)\subseteq\mathcal{C}(j)$. Event $\cup_Z\mathcal{Z}(j,Z)$ is the complement of $\mathcal{C}(j)$. For each Z, event

$\mathcal{Z}(j,Z) = \mathcal{Y}(j,Z) \cup \mathcal{X}(j,Z)$; Event $\mathcal{X}(j,Z)$ (respectively $\mathcal{Y}(j,Z)$) is the event that B is contained inside (respectively, outside) the central ball of a cluster whose radius is $\approx (8/5)^j$.

For each cluster we define the depth of its local density change as a function of the ratio between its radius and its parent reset radius. The parent reset cluster $Y_i(x)$ of a cluster $F_i(x)$ is defined as follows. For any $i > 0$ if $F_i(x) \in \mathcal{R}$ let $Y_i(x) = F_i(x)$, otherwise let $Y_i(x) \in \mathcal{R}$ such that $F_i(x) \in \mathcal{G}_{Y_i(x)}$. The depth of the local density change is defined as

Definition 6. *Let*

$$\alpha_i(x) = \frac{1}{f(\log(2\mathrm{rad}(Y_i(x))/\mathrm{rad}(F_i(x))))}$$

notice that α is a uniform function over $\mathcal{F}$, i.e. *if $u, v \in F_i(x)$ then $\alpha_i(u) = \alpha_i(v)$.*

Given this parameter we define the local density of a node x in a subgraph Y_0 as

Definition 7. *Let*

$$\rho_{Y_0}(x,i) = \frac{|Y_0|}{|B_{Y_0}(x, \alpha_i(x)\mathrm{rad}(F_i(x))/64)|} .$$

We shall use the following Lemma from [2]

Lemma 13. *Let (X,d) be a metric space and $Z \subseteq X$. let $\chi \geq 2$ be a parameter. Given $0 < \Delta < \mathrm{diam}(Z)$ and a center point $v \in Z$, there exists a probability distribution over partitions $(S, \bar{S})$ of Z such that $S = B_{(Z,d)}(v,r)$, and r is chosen from a probability distribution in the interval $[\Delta/4, \Delta/2]$, such that for any $\theta \in (0,1)$ satisfying $\theta \geq \chi^{-1}$, let $\eta = \frac{1}{16}\ln(1/\theta)/\ln\chi$ then for any $x \in Z$, the following holds:*

$$\Pr[B_Z(x,\eta\Delta) \bowtie (S,\bar{S})] \leq \\ (1-\theta)\left[\Pr[B_Z(x,\eta\Delta) \nsubseteq \bar{S}] + 2\chi^{-2}\right] ..$$

Given this lemma we prove a variant of the Uniform Padding Lemma of [2] that is tailored to the construction of our algorithm. There are three main differences. The first difference is that instead of cutting balls we cut cones, the second difference is that the parameter of the cut is defined in a subtle way with respect to the last reset cluster: the local density change of a node is defined as $\rho_{Y_0}(x,i)$ which depends on $\alpha_i(x)$ which depends on $\mathrm{rad}(Y_i(x))/\mathrm{rad}(F_i(x))$ and Equation 5. The final difference is that the hierarchical scheme ensures the relation $\frac{|F_i(x)|}{|Y_i(x)|} \leq c\frac{\mathrm{rad}(F_i(x))}{\mathrm{rad}(Y_i(x))}$.

Lemma 14. *For all $Y_0 \subset X$, $x, y \in G_\epsilon$ and $j > 0$ such that $d(x,y) \leq (\frac{8}{5})^j/(32f(\log(2c/\epsilon)))$:*

$$\Pr[\mathcal{C}(x,y,j) \mid \mathcal{Y}(x,y,j,Y_0)] \leq \frac{2^8 d(x,y)\cdot f(\log(2c/\epsilon))}{(8/5)^j} \cdot \ln\left(\frac{|Y_0|}{|B_{Y_0}(x,(8/5)^j/(64f(\log(2c/\epsilon))))|}\right) .$$

Proof. Fix $Y_0 \subset X$, $x, y \in G_\epsilon$ and $j > 0$, such that $d(x,y) \leq (8/5)^j/(32f(\log(2c/\epsilon)))$. Let $\mathcal{F}^{(i)}$ be any partial laminar family consistent with the event $\mathcal{Y}(x,y,j,Y_0)$, hence $i = i_j(\mathcal{F}^{(i)})$.

Now we bound the probability that $B(x,d(x,y)) \nsubseteq F_{i+1}(x)$ given $\mathcal{F}^{(i)}$ and that the central ball X_0 is disjoint from $B(x,d(x,y))$.

From $B(x,d(x,y)) \subseteq F_i(x)$ follows $|F_i(x)| \geq \epsilon n$. We know by the construction and definition of reset clusters that $\frac{|F_i(x)|}{|Y_i(x)|} \leq c\frac{\mathrm{rad}(F_i(x))}{\mathrm{rad}(Y_i(x))}$ hence $2\mathrm{rad}(Y_i)/\mathrm{rad}(F_i(x)) \leq 2c/\epsilon$ which implies that $\alpha_i(x) \geq \frac{1}{f(\log(2c/\epsilon))}$.

Let $\Delta = \alpha_i(x)\mathrm{rad}(F_i(x))/4$. For $k \geq 1$ let v_k, x_k, $\hat{\chi}_k$ and χ_k be as in the algorithm, and let ℓ_k be the appropriate cone-metric.

Let $\delta = \delta_{x,y,j,Y_0} = \exp\left\{-\frac{2^8 d(x,y)\cdot f(\log(2c/\epsilon))}{(8/5)^j} \cdot \ln\left(\frac{|Y_0|}{|B_{Y_0}(x,(8/5)^j/(64f(\log(2c/\epsilon))))|}\right)\right\}$. If $\delta < e^{-1}$ then the claim is trivial (probability is always bounded by 1), so the interesting cases are when $\delta \geq e^{-1}$. Let $\theta = \delta^{1/2}$. Note that $\theta \geq 2\chi_k^{-1}$ as required (the algorithm actually apllies Lemma 13 on (Y_k, ℓ_k) with x_k as center and the parameter χ_k).

First consider the case that $\rho_{Y_0}(x,i) < 2$, then we claim that $B_{Y_0}(x,d(x,y))$ cannot be cut by a cone:

Since v_1 was chosen as to minimize $\rho_{Y_0}(z,i)$ then $\rho_{Y_0}(v_1,i) < 2$ as well. It implies that both $|B_{Y_0}(v_1,\Delta/16)|, |B_{Y_0}(x,\Delta/16)| > |Y_0|/2$, hence $B_{Y_0}(v_1,\Delta/16) \cap B_{Y_0}(x,\Delta/16) \neq \emptyset$, therefore $d(x,v_1) \leq \Delta/8$. Since $d(x,y) \leq \Delta/8$ and $\ell_1(v_1,x_1) = 0$ follows that $B_{Y_0}(x,d(x,y)) \subseteq B_{(Y_0,\ell_1)}(x_1,\Delta/4)$.

Now assume that $\rho_{Y_0}(x,i) \geq 2$. We now claim that for all $x \in Y_{k-1}$, $\eta_k\Delta \geq d(x,y)$. Recall that $\eta_k = 2^{-4}\ln(1/\theta)/\ln\chi_k = 2^{-5}\ln(1/\delta)/\ln\chi_k$, and notice that if $x \in Y_{k-1}$ then $\rho_{Y_0}(x,i) \geq \hat{\chi}_k$. If $\hat{\chi}_k < 4$ then $\chi_k = 4$ and $\log\rho_{Y_0}(x,i)/\log\chi_k \geq 1/2$, otherwise $\chi_k = \hat{\chi}_k$ and $\log\rho_{Y_0}(x,i)/\log\chi_k \geq 1$. Since $\alpha_i(x)\mathrm{rad}(F_i(x)) \geq (8/5)^j/f(\log(2c/\epsilon))$ we get:

$$\eta_k\Delta \geq \frac{2^8 d(x,y) f(\log(2c/\epsilon)) \cdot \log\rho_{Y_0}(x,i)}{2^5(8/5)^j \log\chi_k} \cdot \frac{(8/5)^j}{4f(\log(2c/\epsilon))} \geq d(x,y).$$

It remains to show that if $x \in X_k$ then
$\Pr[B_{Y_0}(x,\eta_k\Delta) \nsubseteq X_k] \leq 1-\delta \leq \frac{2^8 d(x,y)\cdot f(\log(2c/\epsilon))}{(8/5)^j} \cdot \ln\left(\frac{|Y_0|}{|B_{Y_0}(x,(8/5)^j/(64f(\log(2c/\epsilon))))|}\right)$ as required.

Consider the distribution over partitions of Y_0 into cones $X_1, X_2, \ldots X_t$ as defined above. For $1 \leq m \leq t$, define the events:

$$\begin{aligned}
\mathcal{Z}_m &= \{\forall k, 1 \leq k < m,\ B_{Y_0}(x,\eta_k\Delta) \subseteq Y_k\}, \\
\mathcal{E}_m &= \{\exists k,\ m \leq k < t \text{ s.t. } B_{Y_0}(x,\eta_k\Delta) \bowtie (X_k,Y_k) | \mathcal{Z}_m\}.
\end{aligned}$$

We prove the following inductive claim: For every $1 \leq m \leq t$:

$$\Pr[\mathcal{E}_m] \leq (1-\theta)(1+\theta\mathbb{E}[\sum_{k\geq m}\chi_k^{-1}|\mathcal{Z}_m]). \tag{7}$$

The proof is essentially the same as the one in [2].

Note that $\Pr[\mathcal{E}_t] = 0$. Assume the claim holds for $m+1$ and we will prove for m. Define the events:

$$\begin{aligned}
\mathcal{F}_m &= \{B_{Y_0}(x,\eta_m\Delta) \bowtie (X_m,Y_m)|\mathcal{Z}_m\}, \\
\mathcal{G}_m &= \{B_{Y_0}(x,\eta_m\Delta) \subseteq Y_m|\mathcal{Z}_m\} = \{\mathcal{Z}_{m+1}|\mathcal{Z}_m\}.
\end{aligned}$$

First we bound $\Pr[\mathcal{F}_m]$. Assume first a particular choice of the cones $X_1, \ldots X_{m-1}$ such that event $\mathcal{Z}_m$ occurs. Call this specific event $\mathcal{A}$, then given that $\mathcal{A}$ occurred the point v_m is now determined deterministically, and so is the value of χ_m. Now, applying Lemma 13 we get

$$\begin{aligned}
&\Pr[B_{Y_0}(x,\eta_m\Delta) \bowtie (X_m,Y_m)|\mathcal{A}] \leq \\
&\qquad (1-\theta)(\Pr[B_{Y_0}(x,\eta_m\Delta) \nsubseteq Y_m|\mathcal{A}] + \theta\chi_m^{-1}.
\end{aligned}$$

It follows that

$$\Pr[\mathcal{F}_m] \leq (1-\theta)(\Pr[\bar{\mathcal{G}}_m] + \theta\mathbb{E}[\chi_m^{-1}|\mathcal{Z}_m]).$$

Using the induction hypothesis we prove the inductive claim:

$$\begin{aligned}
\Pr[\mathcal{E}_m] &\leq \Pr[\mathcal{F}_m] + \Pr[\mathcal{G}_m]\Pr[\mathcal{E}_{m+1}] \\
&\leq (1-\theta)(\Pr[\bar{\mathcal{G}}_m] + \theta\mathbb{E}[\chi_m^{-1}|\mathcal{Z}_m]) + \\
&\quad \Pr[\mathcal{G}_m]\cdot(1-\theta)(1+\theta\mathbb{E}[\sum_{k\geq m+1}\chi_k^{-1}|\mathcal{Z}_{m+1}]) \\
&\leq (1-\theta)(1+\theta\mathbb{E}[\sum_{k\geq m}\chi_k^{-1}|\mathcal{Z}_m]),
\end{aligned}$$

Now consider a fixed choice of star-partition $\{X_0, \ldots, X_t\}$. Since the radius of every cone is at least $\Delta/4$, and since for every $k \in [t]$, $\ell_k(v_k,x_k) = 0$ we get that $B_{(Y_0,d)}(v_k,\Delta/16) \subseteq B_{(Y_0,\ell_k)}(x_k,\Delta/4) \subseteq X_k$. Therefore if $k \neq k'$ then $B_{(Y_0,d)}(v_k,\Delta/16) \cap B_{(Y_0,d)}(v_{k'},\Delta/16) = \emptyset$. Hence, we get:

$$\sum_{k\geq m}\chi_k^{-1} \leq \sum_{k\geq m}\hat{\chi_k}^{-1} = \sum_{k\geq m}\frac{|B_{(Y_0,d)}(v_k,\Delta/16)|}{|Y_0|} \leq 1\,.$$

We conclude that if $x \in X_m$

$$\Pr[B_{Y_0}(x,\eta_m\Delta) \nsubseteq X_m] = \Pr[\mathcal{E}_1] \leq$$
$$(1-\theta)(1+\theta \cdot \mathbb{E}[\sum_{k\geq 1}\chi_k^{-1}]) \leq (1-\theta)(1+\theta) = 1-\delta.$$

Since $\mathcal{Y}(x,y,j,Y_0)$ we have that $B(x,d(x,y)) \subseteq Y_0$, hence indeed $\Pr[B(x,d(x,y)) \nsubseteq X_m] \leq 1-\delta$.

□

We complete the algorithm analysis be proving that the expected distortion is scaling. As with many partition based schemes that use local density, the *core* argument is essentially based on the observation that the series $\sum_{a<i\leq b}\log\frac{|B(x,2^i)|}{|B(x,2^{i-1})|}$ is a telescoping series hence it can be bounded by $\log\frac{|B(x,2^b)|}{|B(x,2^a)|}$. When $|B(x,2^a)| \geq \epsilon n$ and b is large enough then this argument gives the essential $O(\log 1/\epsilon)$ scaling ingredient. The following is a technical generalization of this core idea. The main problem is that the local density change $\rho_{Y_0}(x,i)$ of our algorithm is defined as a function of Y_0, but Y_0 is determined by a probabilistic processes. Hence in order for the core telescoping argument to work we need to delicately combine the various probabilistic events in a hierarchical manner. This is done by induction.

Lemma 15. *For any $x,y \in G_\epsilon$ we have*

$$\mathbb{E}[d_T(x,y)] \leq \tilde{O}(\log^2(1/\epsilon))d(x,y).$$

Proof. For any $\epsilon > 0$ fix some $x,y \in G_\epsilon$. Let ℓ be the smallest integer such that $d(x,y) \leq (8/5)^\ell/(64f(\log(2c/\epsilon)))$, and let $\lceil L = \log_{(8/5)}\operatorname{diam}(X)\rceil$. For ease of notation for any $j > 0$ writing $\mathbb{E}_{Z_j}$ means that the expectation is over clusters Z_j such that $(\frac{8}{5})^j \leq \operatorname{rad}(Z_j) < (8/5)^{j+1}$ that contain $B(x,d(x,y)) \subseteq Z_j$ whose distribution is induced by the hierarchical probabilistic star partition algorithm.

Let $k = 2^4c \cdot \log_{(8/5)}\log(1/\epsilon)$. First we prove by induction on $j \geq \ell+k$ the following claim: For any $m \in [\ell, j-1]$ let $h = \max\{m+1, j-k+1\}$, then for any $Z_j \subseteq X$:

$$\begin{aligned}(8)\qquad &\sum_{m=\ell}^{j-1}\mathbb{E}_{Z_h}\left[\Pr\left[\mathcal{C}(m) \mid \mathcal{Z}(Z_h)\right] \mid \mathcal{Z}(Z_j)\right]\cdot(8/5)^m \\ &\leq \quad 2^{10}c\cdot d(x,y)f(\log(1/\epsilon))\sum_{i=j-k+1}^{j}\mathbb{E}_{Z_i}\left[\ln\left(\frac{|Z_i|}{\epsilon n}\right) \mid \mathcal{Z}(Z_j)\right].\end{aligned}$$

The base cases when $j = \ell+k$ is proved similarly to the induction step and we leave it for the reader.

Assume the claim holds for j and prove for $j+1$. Fix any $Z_{j+1} \subseteq X$, for abbreviation let $B_Z(x) = B_Z(x,(8/5)^j/(64f(\log(2c/\epsilon))))$. Let p_j be the probability that $B(x,d(x,y)) \subseteq X_j$, where X_j is the central ball in the star partition of the cluster Z_{j+1}. Consider first the last element in the summation:

$$\begin{aligned}(9)\qquad &\Pr\left[\mathcal{C}(j) \mid \mathcal{Z}(Z_{j+1})\right]\cdot(8/5)^j \\ &\leq \quad \Pr\left[\mathcal{C}_{\text{ball}}(j) \mid \mathcal{Z}(Z_{j+1})\right](8/5)^j + (1-p_j)\mathbb{E}_{Y_j}\left[\Pr[\mathcal{C}(j) \mid \mathcal{Y}(Y_j)](8/5)^j \mid \mathcal{Z}(Z_{j+1})\right]..\end{aligned}$$

Consider the term $\Pr\left[\mathcal{C}_{\text{ball}}(j) \mid \mathcal{Z}(Z_{j+1})\right]$. We choose the radius r of the central ball to be in the "sparsest" of two disjoint strips around x_0: $(1/2,9/16)\operatorname{rad}(Z_{j+1})$ and $(9/16,10/16)\operatorname{rad}(Z_{j+1})$, hence only one of them can contain more than half of the points in Z_{j+1}, and we will choose r from the other one, which contains less than half of the points.

Moreover, the radius is actually in a sub-strip - *i.e.* in the interval $(1/2,1/2+1/32)\operatorname{rad}(Z_{j+1})$ or in $(1/2+3/32,1/2+1/8)\operatorname{rad}(Z_{j+1})$. Hence if $B(x,\alpha_{i_j}(x)\operatorname{rad}(Z_{j+1})/64)$ intersects one of these sub-strips, it will be fully contained within the appropriate strip (recall that $\alpha \leq 1$), which suggest that if the $B(x,\operatorname{rad}(Z_{j+1})/64)$

can be cut by the central ball, it contains less than half of the points in Z_{j+1}, *i.e.* $\rho_{Z_{j+1}}(x,i_j) \geq 2$. We conclude that

$$
\begin{aligned}
(10) \qquad & \Pr\left[\mathcal{C}_{\mathrm{ball}}(j) \mid \mathcal{Z}(Z_{j+1})\right](8/5)^j \\
\leq\ & \frac{2d(x,y)}{\mathrm{rad}(Z_{j+1})/32}(8/5)^j \\
\leq\ & 2^6 d(x,y)\cdot \rho_{Z_{j+1}}(x,i_j) \\
\leq\ & 2^7 d(x,y)(p_j+(1-p_j))\cdot \ln\left(\frac{|Z_{j+1}|}{|B_{Z_{j+1}}(x)|}\right) \\
\leq\ & 2^9 c\cdot d(x,y) f(\log(1/\epsilon)) \\
& \cdot\left(p_j\cdot \mathbb{E}_{X_j}\left[\ln\left(\frac{|Z_{j+1}|}{|B_{X_j}(x)|}\right) \mid \mathcal{Z}(Z_{j+1})\right] + (1-p_j)\mathbb{E}_{Y_j}\left[\ln\left(\frac{|Z_{j+1}|}{|B_{Y_j}(x)|}\right) \mid \mathcal{Z}(Z_{j+1})\right]\right).
\end{aligned}
$$

In the third inequality we used that $\alpha_{i_j}(x) \geq 1/f(\log(2c/\epsilon))$, and in the last inequality we simply reduced the size of $B_{Z_{j+1}}(x)$ and added expectations.

As for the term $\mathbb{E}_{Y_j}\left[\Pr[\mathcal{C}(j) \mid \mathcal{Y}(Y_j)](8/5)^j \mid \mathcal{Z}(Z_{j+1})\right]$, we apply Lemma 14 which suggests that for any $Y_j \subseteq Z_{j+1}$ it is bounded by $2^9 c\cdot d(x,y)\cdot f(\log(1/\epsilon))\cdot \mathbb{E}_{Y_j}\left[\ln\left(\frac{|Z_{j+1}|}{|B_{Y_j}(x)|}\right) \mid \mathcal{Z}(Z_{j+1})\right]$.

Now consider the reminder of the sum, let $h' = \max\{m+1, j-k+2\}$. Since for any $m \in [\ell, j-1]$, $\mathbb{E}_{Z_h}[\cdot \mid \mathcal{Z}(Z_{j+1})] = \mathbb{E}_{Z_j}(\mathbb{E}_{Z_h}[\cdot \mid \mathcal{Z}(Z_j)] \mid \mathcal{Z}(Z_{j+1}))$ we get that

$$
\begin{aligned}
(11) \qquad & \sum_{m=\ell}^{j-1} \mathbb{E}_{Z_{h'}}\left[\Pr\left[\mathcal{C}(m) \mid \mathcal{Z}(Z_{h'})\right] \mid \mathcal{Z}(Z_{j+1})\right]\cdot(8/5)^m \\
=\ & \mathbb{E}_{Z_j}\left[\sum_{m=\ell}^{j-1}\mathbb{E}_{Z_h}\left[\Pr\left[\mathcal{C}(m) \mid \mathcal{Z}(Z_h)\right] \mid \mathcal{Z}(Z_j)\right]\cdot(8/5)^m \mid \mathcal{Z}(Z_{j+1})\right] \\
=\ & p_j\cdot \mathbb{E}_{X_j}\left[\sum_{m=\ell}^{j-1}\mathbb{E}_{Z_h}\left[\Pr\left[\mathcal{C}(m) \mid \mathcal{Z}(Z_h)\right] \mid \mathcal{X}(X_j)\right]\cdot(8/5)^m \mid \mathcal{Z}(Z_{j+1})\right] \\
& +(1-p_j)\cdot \mathbb{E}_{Y_j}\left[\sum_{m=\ell}^{j-1}\mathbb{E}_{Z_h}\left[\Pr\left[\mathcal{C}(m) \mid \mathcal{Z}(Z_h)\right] \mid \mathcal{Y}(Y_j)\right]\cdot(8/5)^m \mid \mathcal{Z}(Z_{j+1})\right]
\end{aligned}
$$

Notice that h' was changed to h, meaning that we added expectation over level $j-k+1$ as well, this does not change the value of the expression. Applying the induction hypothesis to Equation 11 yields

$$
\begin{aligned}
(12) \qquad & \sum_{m=\ell}^{j-1} \mathbb{E}_{Z_{h'}}\left[\Pr\left[\mathcal{C}(m) \mid \mathcal{Z}(Z_{h'})\right] \mid \mathcal{Z}(Z_{j+1})\right]\cdot(8/5)^m \\
\leq\ & 2^{10} c\cdot d(x,y) f(\log(1/\epsilon)) p_j \cdot \mathbb{E}_{X_j}\left[\sum_{i=j-k+1}^{j}\mathbb{E}_{Z_i}\left[\ln\left(\frac{|Z_i|}{\epsilon n}\right) \mid \mathcal{X}(X_j)\right] \mid \mathcal{Z}(Z_{j+1})\right] \\
& +2^{10} c\cdot d(x,y) f(\log(1/\epsilon))(1-p_j)\mathbb{E}_{Y_j}\left[\sum_{i=j-k+1}^{j}\mathbb{E}_{Z_i}\left[\ln\left(\frac{|Z_i|}{\epsilon n}\right) \mid \mathcal{Y}(Y_j)\right] \mid \mathcal{Z}(Z_{j+1})\right]
\end{aligned}
$$

We now have all the ingredients to prove the inductive claim of Equation 8. For abbreviation let $W = 2^9 c\cdot d(x,y) f(\log(1/\epsilon))$.

$$
\begin{aligned}
&\sum_{m=\ell}^{j} \mathbb{E}_{Z_{h'}}\left[\Pr\left[\mathcal{C}(m) \mid \mathcal{Z}(Z_{h'})\right] \mid \mathcal{Z}(Z_{j+1})\right] \cdot (8/5)^m \\
&\leq \quad W \cdot \left(p_j \cdot \mathbb{E}_{X_j}\left[\ln\left(\frac{|Z_{j+1}|}{|B_{X_j}(x)|}\right) \mid \mathcal{Z}(Z_{j+1})\right] + (1-p_j)\mathbb{E}_{Y_j}\left[\ln\left(\frac{|Z_{j+1}|}{|B_{Y_j}(x)|}\right) \mid \mathcal{Z}(Z_{j+1})\right]\right) \\
&\quad\quad + W \cdot (1-p_j)\mathbb{E}_{Y_j}\left[\ln\left(\frac{|Z_{j+1}|}{|B_{Y_j}(x)|}\right) \mid \mathcal{Z}(Z_{j+1})\right] \\
&\quad\quad + 2W \cdot \left(p_j \cdot \mathbb{E}_{X_j}\left[\sum_{i=j-k+1}^{j} \mathbb{E}_{Z_i}\left[\ln\left(\frac{|Z_i|}{\epsilon n}\right) \mid \mathcal{X}(X_j)\right] \mid \mathcal{Z}(Z_{j+1})\right]\right) \\
&\quad\quad + 2W \cdot \left((1-p_j)\mathbb{E}_{Y_j}\left[\sum_{i=j-k+1}^{j} \mathbb{E}_{Z_i}\left[\ln\left(\frac{|Z_i|}{\epsilon n}\right) \mid \mathcal{Y}(Y_j)\right] \mid \mathcal{Z}(Z_{j+1})\right]\right) \\
&= \quad W \cdot p_j \cdot \mathbb{E}_{X_j}\left[\ln\left(\frac{|Z_{j+1}|}{|B_{X_j}(x)|}\right) + \sum_{i=j-k+1}^{j} \mathbb{E}_{Z_i}\left[\ln\left(\frac{|Z_i|}{\epsilon n}\right) \mid \mathcal{X}(X_j)\right] \mid \mathcal{Z}(Z_{j+1})\right] \\
&\quad\quad + 2W \cdot (1-p_j)\mathbb{E}_{Y_j}\left[\ln\left(\frac{|Z_{j+1}|}{|B_{Y_j}(x)|}\right) + \sum_{i=j-k+1}^{j} \mathbb{E}_{Z_i}\left[\ln\left(\frac{|Z_i|}{\epsilon n}\right) \mid \mathcal{Y}(Y_j)\right] \mid \mathcal{Z}(Z_{j+1})\right] \\
&\quad\quad + W \cdot p_j \cdot \mathbb{E}_{X_j}\left[\sum_{i=j-k+1}^{j} \mathbb{E}_{Z_i}\left[\ln\left(\frac{|Z_i|}{\epsilon n}\right) \mid \mathcal{X}(X_j)\right] \mid \mathcal{Z}(Z_{j+1})\right] \\
&\leq \quad W \cdot p_j \sum_{i=j-k+2}^{j+1} \mathbb{E}_{Z_i}\left[\ln\left(\frac{|Z_i|}{\epsilon n}\right) \mid \mathcal{Z}(Z_{j+1})\right] + 2W \cdot (1-p_j) \sum_{i=j-k+2}^{j+1} \mathbb{E}_{Z_i}\left[\ln\left(\frac{|Z_i|}{\epsilon n}\right) \mid \mathcal{Z}(Z_{j+1})\right] \\
&\quad\quad + W \cdot p_j \sum_{i=j-k+2}^{j+1} \mathbb{E}_{Z_i}\left[\ln\left(\frac{|Z_i|}{\epsilon n}\right) \mid \mathcal{Z}(Z_{j+1})\right] \\
&\leq \quad 2^{10} c \cdot d(x,y) f(\log(1/\epsilon)) \sum_{i=j-k+2}^{j+1} \mathbb{E}_{Z_i}\left[\ln\left(\frac{|Z_i|}{\epsilon n}\right) \mid \mathcal{Z}(Z_{j+1})\right].
\end{aligned}
$$

The first inequality follows from Equation 9, Equation 10 and Equation 12. The second equality is just a re-ordering of terms. The third inequality is the telescope argument, it holds since for any choice of $X_j \subseteq Z_{j+1}$, and any choice of $Z_{j-k+1} \subseteq X_j$ by definition $\mathrm{rad}(Z_{j-k+1}) \leq (5/8)^{k-1}\mathrm{rad}(X_j) \leq \frac{\mathrm{rad}(X_j)}{2^7 f(\log(2c/\epsilon))}$, since $x \in Z_{j-k+1}$ follows $Z_{j-k+1} \subseteq B_{X_j}(x)$. The argument for Y_j is similar. So the elements depending on X_j and Y_j cancel out, and we don't need the expectation on X_j and Y_j anymore.

Let $j = L+1$, $Z_j = X$, then applying Lemma 12 and Equation 8 completes the proof.

$$\begin{aligned}
\mathbb{E}[d_T(x,y)] &\leq \sum_{m=1}^{L} \Pr[\mathcal{C}_m] \cdot 2\mathrm{rad}(T[F_{i_m}(x)]) \\
&\leq 4c' \sum_{m=1}^{\ell-1} (8/5)^m + 4c' \sum_{m=\ell}^{L} \mathbb{E}_{Z_h}\left[\Pr[\mathcal{C}_m \mid \mathcal{Z}(Z_h)] \mid \mathcal{Z}(X)\right](8/5)^m \\
&\leq 4c'(5/3)(8/5)^\ell + 2^{12} c \cdot c' \cdot d(x,y) f(\log(1/\epsilon)) \sum_{i=L-k+1}^{L} \mathbb{E}_{Z_i}\left[\ln\left(\frac{|Z_i|}{\epsilon n}\right)\right] \\
&\leq 8c' \cdot d(x,y) 64 f(\log(2c/\epsilon)) + 2^{12} c \cdot c' \cdot d(x,y) f(\log(1/\epsilon)) 2^4 c \cdot \log\log(1/\epsilon) \ln(n/(\epsilon n)) \\
&= \tilde{O}\left(\log^2(1/\epsilon)\right)
\end{aligned}$$

□

References

[1] Ittai Abraham, Yair Bartal, Hubert T.-H. Chan, Kedar Dhamdhere, Anupam Gupta, Jon M. Kleinberg, Ofer Neiman, and Aleksandrs Slivkins. Metric embeddings with relaxed guarantees. In *FOCS*, pages 83–100. IEEE Computer Society, 2005.

[2] Ittai Abraham, Yair Bartal, and Ofer Neiman. Advances in metric embedding theory. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 271–286, New York, NY, USA, 2006. ACM Press.

[3] Noga Alon, Richard M. Karp, David Peleg, and Douglas West. A graph-theoretic game and its application to the k-server problem. *SIAM J. Comput.*, 24(1):78–100, 1995.

[4] Vassilis Athitsos and Stan Sclaroff. Database indexing methods for 3d hand pose estimation. In *Gesture Workshop*, pages 288–299, 2003.

[5] Y. Bartal, B. Bollobás, and M. Mendel. Ramsey-type theorems for metric spaces with applications to online problems. *Journal of Computer and System Sciences*, 72(5):890–921, August 2006. Special Issue on FOCS 2001.

[6] Yair Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *37th Annual Symposium on Foundations of Computer Science (Burlington, VT, 1996)*, pages 184–193. IEEE Comput. Soc. Press, Los Alamitos, CA, 1996.

[7] Yair Bartal. On approximating arbitrary metrics by tree metrics. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 183–193, 1998.

[8] Yair Bartal. Graph decomposition lemmas and their role in metric embedding methods. In *12th Annual European Symposium on Algorithms*, pages 89–97, 2004.

[9] Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. On metric ramsey-type phenomena. *Annals Math*, 162(2):643–709, 2005.

[10] Yair Bartal and Manor Mendel. On low dimensional Lipschitz embeddings of ultrametrics, 2002. Manuscript.

[11] Yair Bartal and Manor Mendel. Multi-embedding and path approximation of metric spaces. In *SODA '03: Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 424–433, Philadelphia, PA, USA, 2003. Society for Industrial and Applied Mathematics.

[12] Michael Elkin, Yuval Emek, Daniel A. Spielman, and Shang-Hua Teng. Lower-stretch spanning trees. In *STOC '05: Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 494–503, New York, NY, USA, 2005. ACM Press.

[13] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *STOC '03: Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 448–455. ACM Press, 2003.

[14] Eran Halperin, Jeremy Buhler, Richard M. Karp, Robert Krauthgamer, and B. Westover. Detecting protein sequence conservation via metric embeddings. In *ISMB (Supplement of Bioinformatics)*, pages 122–129, 2003.

[15] Sariel Har-Peled and Manor Mendel. Fast construction of nets in low-dimensional metrics and their applications. *SIAM J. Comput*, 35(5):1148–1184, 2006.

[16] G. Hjaltason and H. Samet. Contractive embedding methods for similarity searching in metric spaces, 2000.

[17] Gabriela Hristescu and Martin Farach-Colton. Cofe: A scalable method for feature extraction from complex objects. In *DaWaK*, pages 358–371, 2000.

[18] P. Indyk. Algorithmic applications of low-distortion geometric embeddings. In *Proceedings of the 42nd Annual Symposium on Foundations of Computer Science*, pages 10–33, 2001.

[19] Jon M. Kleinberg, Aleksandrs Slivkins, and Tom Wexler. Triangulation and embedding using small sets of beacons. In *FOCS*, pages 444–453, 2004.

[20] Yuri Rabinovich and Ran Raz. Lower bounds on the distortion of embedding finite metric spaces in graphs. *Discrete & Computational Geometry*, 19(1):79–94, 1998.

[21] Yuval Shavitt and Tomer Tankel. Big-bang simulation for embedding network distances in euclidean space. *IEEE/ACM Trans. Netw.*, 12(6):993–1006, 2004.

[22] Liying Tang and Mark Crovella. Geometric exploration of the landmark selection problem. In *PAM*, pages 63–72, 2004.

[23] Bang Ye Wu, Giuseppe Lancia, Vineet Bafna, Kun-Mao Chao, R. Ravi, and Chuan Yi Tang. A polynomial time approximation scheme for minimum routing cost spanning trees. In *SODA '98: Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms*, pages 21–32. Society for Industrial and Applied Mathematics, 1998.